BULLETIN

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

EARLE R. HEDRICK

DAVID R. CURTISS WILLIAM R. LONGLEY

WITH THE ASSISTANCE OF

HAROLD T. DAVIS HARRY W. KUHN

THEOPHIL H. HILDEBRANDT DERRICK N. LEHMER

JOHN R. KLINE HENRY L. RIETZ

CAROLINE E. SEELY

VOLUME XXXIX, NUMBER 2

FEBRUARY, 1933

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1933

PUBLISHED MONTHLY

Whole No. 401 $9.00 a Year

Entered as second class matter April 24, 1926 at the post office at Menasha, Wis., under the act of August 24, 1912.

AMERICAN MATHEMATICAL SOCIETY

Mail Address: 501 West 116th Street, New York, N.Y., for the Secretary, Treasurer, Librarian, and Clerk.

Office: Room 413, Physics Building, Columbia University, New York City.

President: Professor A. B. Coble, University of Illinois, Urbana, Ill.

Secretary: Dean R. G. D. Richardson, Brown University, Providence, R.I.

Associate Secretaries: Professor M. H. Ingraham, University of Wisconsin, Madison, Wis.; Professor J. R. Kline, University of Pennsylvania, Philadelphia, Penna.; Dean T. M. Putnam, University of California, Berkeley, Calif.

Treasurer: Professor G. W. Mullins, Barnard College in Columbia University, New York City.

Librarian: Professor R. C. Archibald, Brown University, Providence, R.I.

Clerk: Dr. Caroline E. Seely, 413 Physics Building, Columbia University, New York City.

Editor-in-Chief of the Bulletin: Professor E. R. Hedrick, University of California at Los Angeles, Los Angeles, Calif.

Managing Editor of the Transactions: Professor R. D. Carmichael, University of Illinois, Urbana, Ill.

Chairman, Colloquium Editorial Committee: Professor R. L. Moore, University of Texas, Austin, Tex.

Managing Editor of the American Journal of Mathematics: Professor Abraham Cohen, Johns Hopkins University, Baltimore, Md.

Representatives of the Society in the Division of Physical Sciences of the National Research Council: Professors A. B. Coble, E. R. Hedrick, and D. R. Curtiss.

Representatives on the Council of the American Association for the Advancement of Science: Professors Louis Ingold and M. H. Ingraham.

Chairman, Committee on Arrangements for the Semi-Centennial: Professor T. S. Fiske, Columbia University, New York City.

THERMODYNAMICS AND RELATIVITY*

BY R. C. TOLMAN

1. *Introduction.* We have met to do honor to the memory of Josiah Willard Gibbs. By the labors of this master, the classical principles of thermodynamics were given their most complete and comprehensive expression. As the subject for the tenth memorial lecture, it seems appropriate to discuss the extensions to these classical principles which have since been made necessary by Einstein's discovery of the special and general theories of relativity.

The need for an extension of thermodynamics to relativity arises in two ways. In the first place, the classical thermodynamics was—perhaps unintentionally but nevertheless actually—only developed for systems which were tacitly assumed to be at rest with respect to the observer, and further investigation is necessary for the treatment of thermodynamic systems which are moving relative to the spatial coordinates in use. This further investigation must be carried out with the help of those principles for the inter-comparison of measurements—made by observers in uniform relative motion to each other—which form the subject matter of the special theory of relativity.

In the second place, the older thermodynamics tacitly assumed that the behavior of thermodynamic systems could be described with the help of ideas as to the nature of space and time which we now know to be approximately valid only for a limited range of space-time and in the absence of strong gravitational fields. The considerations of the classical thermodynamics were thus actually limited to the treatment of small enough systems and weak enough gravitational fields so that the deviations from this kind of space-time could be neglected, and the Newtonian theory of gravitation could be applied as a close enough approximation. In order, however, to investigate the thermodynamic behavior of large portions of the universe as we may

* The Tenth Josiah Willard Gibbs Lecture, delivered at Atlantic City, December 29, 1932, under the auspices of the American Mathematical Society, at a joint meeting of the Society with the American Physical Society, and Section A of the American Association for the Advancement of Science.

wish to do in connection with cosmological problems, and in order to obtain even in the case of small systems more precise expressions for the thermodynamic effects of gravity, it becomes necessary to extend thermodynamics to general relativity, and to make use of the more valid ideas as to the nature of space and time and the more precise theory of gravitation which Einstein has now provided.

2. *The Character and Validity of Thermodynamics and Relativity.* In carrying out these proposed extensions of thermodynamics to relativity, it proves possible to combine the known principles of thermodynamics with those of special and general relativity in a very natural manner with only small and apparently rational additions in the way of new hypothesis. Hence the character and validity of the system of relativistic thermodynamics that we obtain is largely dependent on the character and validity of the two component sciences.

In character, the classical thermodynamics may be regarded as a macroscopic, phenomenological science, which has no actual need for that interesting kind of support that can be furnished by the microscopic atomic considerations of statistical mechanics, but which attempts to treat the gross behavior of matter with the help of those generalized descriptions of the results of numerous gross experiments on the mechanical equivalent of heat and on the efficiency of heat engines, which we call the first and second laws of thermodynamics.

As to the validity of thermodynamics, we have feelings of great confidence, on account of the extensive experimental verification which exists, not only directly for the two laws themselves, but for an extraordinary number of consequences which have been drawn from them—often by elaborate but logical trains of deductive reasoning. Further additions to the principles of thermodynamics may be found, such as the newer so-called third law of Nernst and Planck, but we cannot escape the conviction that, so long as the human mind retains its present ideas of rationality, these additions are likely to prove—as in the case mentioned—supplementary rather than destructive.

Relativity, a science which changes as it does our very ideas as to the nature of space and time, has much more fundamental and far-reaching implications than thermodynamics and cannot

be so easily characterized. There are, however, certain similarities between the two sciences which may be emphasized.

In the first place, at least in its present stage of development, relativity must also be regarded as a macroscopic theory dealing with ideas as to the nature of space and time which have been directly derived from macroscopic experiences. Indeed, in view of Heisenberg's uncertainty principle and the great difficulties which have been encountered in all attempts to construct a satisfactory relativistic quantum mechanics, we may even doubt whether these ideas as to space and time are really suitable for microscopic considerations. This, however, offers no difficulties if we are to combine with another macroscopic science such as thermodynamics.

In the second place, although we are often inclined to be specially impressed by the wonderful conceptual content of the theory of relativity, we may here emphasize its not negligible character as a phenomenological or descriptive science.

Thus the first of the two postulates of the special theory of relativity may be regarded as a generalized description of many failures to detect the absolute velocity of the earth's motion. And the second postulate may be regarded as a mere empirical statement of that constancy in the velocity of light, which is specially clearly demonstrated in the case of distant double stars by the lack of any effect from the changing motions of the members of the doublet on the time needed for their light to reach the earth.

Turning, moreover, to the two postulates necessary for the general theory of relativity, the principle of equivalence may be regarded not unfairly as a reasonably generalized description of Galileo's discovery that all bodies fall at the same rate. The principle of covariance, however, is on a somewhat different footing, since as first pointed out by Kretschmann—given sufficient mathematical ingenuity—any physical law whatever could undoubtedly be expressed in covariant language the same for all coordinate systems, so that the principle of covariance can imply no necessary physical consequences. Nevertheless, as emphasized by Einstein, the actual phenomena of physics must themselves be independent of the choice of coordinate system, since this is a conceptual introduction on the part of the scientist which may be made in any way that may suit his conveni-

ence or please his fancy. Hence the actual employment of invariant forms of expression in searching for the appropriate axioms of physics is desirable in order to avoid the introduction of unsuspected assumptions which might otherwise be insinuated by the use of special coordinates. We can then also, somewhat facetiously, emphasize the phenomenological character of the principle of covariance, by regarding it as a generalized description of the familiar phenomenon, that the purely conceptual activities of man—in inventing imaginary coordinate systems—are likely in first approximation to have no immediate effect on the laws of physics.

As to the validity of the theory of relativity, we have to rely on three different kinds of evidence.

In the first place, we may put its agreement with a great range of diverse facts from different branches of science, which as isolated phenomena can often be attractively explained in terms of pre-relativistic notions, but which as a whole have only been successfully correlated with the help of the theory of relativity.

In the second place, we must put those special observations which distinguish as uniquely as may be between the predictions of relativity and those which would result from other points of view. Here we have in the case of the first postulate of special relativity the demonstration of the Lorentz contraction by the Michelson-Morley experiment and all its now numerous repetitions, if we may include the extensive work of Professor Miller as demonstrating this contraction at least as the primary effect. And we also have the remarkable demonstration of Einstein's time dilation by the beautiful experiments of Kennedy and Thorndike. In the case of the second postulate of the special theory, we have as most important the precise analysis of double star orbits by de Sitter. And turning to the general theory of relativity, we have the entirely satisfactory results of the three crucial tests provided by the rotation of the perihelion of Mercury, the bending of light in passing the sun, and the shift in the wave-length of light originating on the surface of the sun and on that of the companion to Sirius.

Finally, as a third kind of evidence for judging the validity of relativity, we must not neglect the bearings of that wonderful internal coherence of the theory, with its simple foundation and elaborate but logical superstructure, which so well attests the

genius of Einstein. Although such qualities can of themselves provide no guarantee as to correspondence with external phenomena, we can, nevertheless, regard them as indicating that such correspondence—when found for our present limited range of observation—is likely to persist over a much wider range of possible experience.

Like all parts of science, the theory of relativity will presumably be subject to future modifications and additions, such, for example, as might be provided by a successful unified field theory. Nevertheless, just as the Einstein theory has retained the Newtonian theory of gravitation as an exceedingly satisfactory first approximation, we may expect at least for a long time that such changes will here—as well as in the case of thermodynamics—be supplementary rather than destructive.

The character of the two sciences of thermodynamics and relativity, which we are going to combine, is thus sufficiently similar so that we may have no hesitations on that score, and may expect the resulting relativistic thermodynamics to be itself a macroscopic theory suitable for use in the description of the gross phenomena of the external world. And the validity of the two component sciences is sufficiently established so that for the present we may concentrate attention, as we shall in what follows, on the rationality of that small amount of additional hypothesis which we must introduce to effect the combination.

3. *The Extension of Thermodynamics to Special Relativity.* We are now ready to consider the actual procedure adopted in the extension of thermodynamics, first to special relativity and then to general relativity. The extension to special relativity, so as to obtain a suitable thermodynamic theory for moving systems, was made by Planck* and by Einstein,† as early as the year 1907, in that brilliant period of development which was initiated by Einstein's publication of the elements of special relativity only two years previous.

a. *Special Relativity and the First Law of Thermodynamics.* In order to appreciate the nature of this extension, let us begin by

* Planck, Berlin Berichte, 1907, p. 542; Annalen der Physik, vol. 26 (1908), p. 1.

† Einstein, Jahrbuch der Radioaktivität und Electronik, vol. 4 (1907), p. 411.

seeing what happens to the first law of thermodynamics when the extension is made.

In the classical thermodynamics for systems at rest with respect to the observer, we have found it important to distinguish two ways in which there can be an interchange of energy between a system and its surroundings, namely, through the flow of heat into the system from its surroundings and through the performance of work by the system on its surroundings. Making use of this distinction, and making use of the principle of the conservation of energy, which requires that any alteration in energy content can only result from interchange with the surroundings, we then write the first law of thermodynamics in the form given by equation (1)

$$\Delta E = Q - W, \tag{1}$$

where ΔE is the increase in the energy of the system which accompanies the influx in heat Q and the performance of work W against external forces.

In *form* this equation can be taken over without modification into the thermodynamics of moving systems, in the first place, since the special theory of relativity has done nothing to upset the principle of the conservation of energy, and in the second place, since we shall still wish to distinguish between the energy transfer W corresponding to work done against macroscopic external forces and the other modes of transfer which we call the flow of heat Q.

In the application of this equation to moving instead of stationary systems, however, an important difference—which would not have been suspected in prerelativistic days—now arises on account of the relations between mass, energy, and momentum made clear by Einstein's work. To illustrate this difference, let us consider—as we usually do in thermodynamics—only very simple systems consisting of a given amount of thermodynamic fluid or working substance which exerts a pressure on its surroundings.

If such a system is *at rest*, the only way it can do work on its surroundings is by a change in volume under this pressure, and the application of the first law equation (1) then gives us simply

$$dE_0 = dQ_0 - p_0 dv_0, \tag{2}$$

where the subscript $(_0)$ has been added to indicate that the quantities involved are all referred to coordinates in which the system is at rest.

If such a system is *in motion*, however, its momentum will in general change with its energy content even though we hold the velocity constant, owing to the special relativity relation which associates mass with energy. Hence in applying the first law equation to moving systems, even in the simple case of constant velocity, we shall have to include, in addition to the work done against external pressure, the work done against the external force involved in the change in momentum. We must then write in general

$$dE = dQ - p\,dv + \bar{u}\cdot d\bar{G}, \tag{3}$$

where the last term is the scalar product of the velocity of the system $\bar{u}$ and its change in momentum $d\bar{G}$. Moreover, in making use of this equation, we must employ the special relativity relation connecting the momentum of the system with its energy flow

$$\bar{G} = \frac{E + pv}{c^2}\,\bar{u}, \tag{4}$$

where c is the velocity of light, and the term $E\bar{u}/c^2$ gives the momentum due to the transport of the energy of the system as a whole, and the term $p\,v\,\bar{u}/c^2$ corresponds to the additional flow of energy resulting from the work done on the moving volume by the action of the external pressure.

With the help of these two expressions for the first law (3), and for the momentum of a moving system (4), we can then obtain transformation equations which will give us expressions for all the quantities involved in the application of the first law to moving systems, in terms of the analogous quantities as they would be measured by a local observer moving with the system. In accordance with the known equations for force, and the Lorentz contraction for moving volumes, we are already provided by the special theory of relativity with the simple transformations for pressure and volume

$$p = p_0, \qquad v = v_0(1 - u^2/c^2)^{1/2}. \tag{5}$$

Furthermore, considering first an adiabatic acceleration in which the velocity of our system is changed without flow of heat or change in internal condition as measured by a local observer, and then considering more general processes in which flow of heat is permitted, we readily obtain as the transformation equations for energy and heat the two expressions

$$E = \frac{E_0 + p_0 v_0 \dfrac{u^2}{c^2}}{(1 - u^2/c^2)^{1/2}}, \quad dQ = dQ_0(1 - u^2/c^2)^{1/2}. \tag{6}$$

This gives all the apparatus necessary for the application of the first law of thermodynamics to moving systems. It is to be specially noted that so far no new assumptions, beyond those already present in the mechanics of special relativity, have been introduced into our system of thermodynamics, except, if you wish, our procedure in still giving the name heat to that part of the energy transfer which does not take place through the work done against macroscopic external forces.

b. *Special Relativity and the Second Law of Thermodynamics.* Let us now turn to the more characteristically thermodynamic considerations involved in the application of the second law of thermodynamics, and examine the fate of this principle when the extension to special relativity is made.

In the classical thermodynamics the full content of the second law could be conveniently condensed into the very simple expression

$$\Delta S \geqq \int \frac{dQ}{T}, \tag{7}$$

where the left-hand side gives the increase in the entropy content S when a system changes from one state to another, and the right-hand side is to be obtained by dividing each element of heat dQ absorbed by its temperature T, and summing up for the whole process by which the system changes from its initial to its final state.

The sign of equality ($=$) in this expression applies to reversible processes which take place with that highest possible efficiency, which would just be sufficient to permit a return *both* of the system and its surroundings to their original state. And

the sign of inequality ($>$) applies to those less efficient, actual processes which we ordinarily encounter in nature. With the help of the relation of equality we can then calculate the entropy of any system by considering an ideal reversible process by which it could be brought from its standard state to the state under consideration. And with the help of the two relations of equality and inequality, we codify all that extraordinary range of information as to the equilibrium and efficiency of physical-chemical processes which is subservient to the second law.

In making the extension to special relativity, it was found possible to take over this expression for the second law of thermodynamics as a postulate without any change at all in form. And this was evidently a rational thing to try to do since it preserves the constancy of entropy for purely mechanical processes, makes the increase of entropy for reversible thermal processes dependent on the transfer of energy in forms other than work, and retains with the help of the sign of inequality those opportunities for irreversibility and spontaneous increase in entropy which lie at the heart of thermodynamic considerations.

In applying this expression to moving systems we must of course substitute values for entropy, heat, and temperature which are appropriate for a moving system, and hence we shall desire transformation equations which will permit us to calculate these quantities in terms of the analogous quantities which would be directly measured by a local observer travelling with the system in question.

In the case of heat, we are already provided by the application of the first law with the transformation equation

$$dQ = dQ_0(1 - u^2/c^2)^{1/2}. \tag{8}$$

In the case of entropy, we are then directly led by the postulate itself to the conclusion that the entropy of a system must be an invariant for the Lorentz transformation

$$S = S_0 \tag{9}$$

owing to the possibility of changing the velocity of a system by a quasi-static reversible adiabatic acceleration, which leaves the internal state and proper entropy S_0 unaltered on account of

the quasi-static character of the acceleration, and leaves the entropy S unaltered on account of the reversible and adiabatic character of the acceleration. This invariance of entropy is, moreover, in evident agreement with the statistical mechanical interpretation which relates the entropy of a system to the probability of its state, a quantity which could hardly be a function of the velocity with which the observer happens to be moving past the system.

Finally in the case of temperature, by combining the requirements of the postulate itself with the two transformation equations already obtained, it is evident that we are now necessarily led to the relation

$$T = T_0(1 - u^2/c^2)^{1/2} \tag{10}$$

in order that the postulated law (7) may apply to the description of a given change in state both from the point of view of a local observer moving with the system and from the point of view of other observers with respect to which the system is in motion.

c. *Discussion of the Extension to Special Relativity.* This completes all that is necessary for the extension of thermodynamics to special relativity.

It will be seen that the additions in the way of new hypotheses, beyond what is already contained in the special theory of relativity and in the classical thermodynamics, have really been very small and apparently rational. Indeed, it seems fair to say that these additions consist solely in the assumption that the second law of thermodynamics, as expressed in the usual well known form given by (7), will not break down when we turn to the consideration of moving systems, and that the quantity dQ occurring in this expression must still be interpreted as that part of the energy transfer which cannot be considered as work done against macroscopic external forces.

It should also be noted that the results which are given by this extended theory are entirely coherent with the accepted body of theoretical physics. For example, the application of this theory to determine the dynamical properties of a moving enclosure filled with black-body radiation leads to the same results as were

originally obtained by Mosengeil* from strictly electromagnetic considerations. And the transformation equation given for heat which we have regarded as derived from an application of the mechanics of special relativity to the behavior of a portion of fluid, agrees with that which can be derived from electromagnetic considerations for the Joule heating effect in a moving electrical conductor. Most important of all, however, it should be noted that the extension has been so devised that any predictions which we make with its help as to the behavior of a given system moving with a *constant* velocity u will completely agree with those which would be made with the help of the classical thermodynamics by a local observer who moves along with the system in question.

It is important to emphasize these qualities of rationality and coherence, since our judgement as to the validity of this extension of thermodynamics must be largely based thereon. Any direct test of the extension would for the present be out of the question, since all the various thermodynamic quantities for moving systems were found to differ from the analogous ones for stationary systems only by terms of the order of u^2/c^2 or higher, and we could only expect differences of this practically undetectable order for any thermodynamic theory of moving systems that might be proposed.

The usefulness of the extension consists partly in the ease with which we can now treat problems by simple thermodynamic methods which would otherwise involve complicated kinetic theory or electromagnetic considerations, as in the case of the moving enclosure filled with radiation. The usefulness of the extension depends mainly, however, on the increased insight which we now have into the nature of thermodynamics and thermodynamic quantities. Thus the invariance to the Lorentz transformation for entropy and for the ratio of heat to temperature provided by the special theory of relativity prove essential for the further extension of thermodynamics to general relativity to which we must now turn.

4. *The Extension of Thermodynamics to General Relativity.* In

* Mosengeil, Annalen der Physik, vol. 22 (1907), p. 867. The results of Mosengeil were employed by Planck in his method of obtaining the extension of thermodynamics to special relativity.

the general theory of relativity, the space-time continuum in which physical events take place is regarded as characterized by the formula for interval

$$(11) \qquad \begin{aligned} ds^2 &= g_{11}dx_1^2 + 2g_{12}dx_1dx_2 + \cdots + g_{44}dx_4^2 \\ &= g_{\mu\nu}dx^\mu dx^\nu, \qquad (g_{\mu\nu} = g_{\nu\mu}), \end{aligned}$$

where x_1, x_2 and x_3 are the three spatial coordinates that are being used, x_4 is the temporal coordinate, and the $g_{\mu\nu}$ are the ten gravitational potentials. The dependence of these gravitational potentials on the distribution of matter and energy is given by Einstein's ten field equations

$$(12) \qquad -8\pi T^{\mu\nu} = R^{\mu\nu} - \tfrac{1}{2}Rg^{\mu\nu} + \Lambda g^{\mu\nu},$$

where $T^{\mu\nu}$ is the energy-momentum tensor, $R^{\mu\nu}$ and R are obtained from the Riemann-Christoffel tensor by contraction, and Λ, the so-called cosmological constant, is a quantity which is observationally known in any case to be exceedingly small when expressed in reciprocal square centimeters, and may well be zero. Finally, the motion of free particles and light rays in this space-time continuum is determined by the equation

$$(13) \qquad \delta \int ds = 0,$$

with ds greater than zero for material particles and equal to zero for light rays.

The results predicted by these fundamental equations of general relativity are in satisfactory agreement with all the facts that are now at our disposal, and, in particular, agree with the astronomical observations which have furnished the three crucial tests of relativity.

In order to include thermodynamics within this framework, we must now enquire into the analogues in general relativity of the ordinary first and second laws of thermodynamics.

a. *The Analogue of the First Law in General Relativity.* In the case of the first law the procedure to be adopted is clear. In the classical thermodynamics the first law was an expression of the principle of the conservation of energy as applied to small stationary systems in the absence of a gravitational field, and in relativistic thermodynamics we must evidently use as the ana-

logue of the first law the more general energy-momentum principle provided by relativistic mechanics.*

This principle can be expressed by the very simple *tensor* equation

$$(T^{\mu\nu})_\nu = 0, \tag{14}$$

and may be regarded as an immediate result of Einstein's field equations (12), since the tensor divergence of the expression there given for the energy-momentum tensor $T^{\mu\nu}$ can be shown to be necessarily identically equal to zero. For purposes of computation it is often more convenient to rewrite this equation in the *tensor density* form

$$\frac{\partial \mathfrak{T}_\mu^\nu}{\partial x^\nu} - \frac{1}{2}\mathfrak{T}^{\alpha\beta}\frac{\partial g_{\alpha\beta}}{\partial x^\mu} = 0. \tag{15}$$

And to obtain an insight into the nature of the principle, it is sometimes useful to rewrite it in the form of an ordinary divergence as expressed by the non-tensor yet nevertheless *covariant* equation

$$\frac{\partial}{\partial x^\nu}(\mathfrak{T}_\mu^\nu + \mathfrak{t}_\mu^\nu) = 0, \tag{16}$$

where the pseudo-tensor density of potential energy and momentum $\mathfrak{t}_\mu^\nu$ is defined for *all systems* of coordinates in such a way that we can substitute $\partial \mathfrak{t}_\mu^\nu/\partial x^\nu$ for the second term of (15).

To remind us of the physical significance of these familiar equations of relativistic mechanics, it will be recalled that the equations reduce in the absence of a gravitational field to the ordinary principles of special relativity for the conservation of the energy and momentum directly associated with matter and radiation. In general, however, in the presence of gravitational fields, it will be evident from the third form (16) in which the equations have been written, that they will lead to conservation laws only when we include—along with the energy and momentum directly associated with matter and radiation—the potential energy and momentum of the gravitational field, which

* Tolman, Proceedings of the National Academy, vol. 14 (1928), p. 268; Physical Review, vol. 35 (1930), p. 875.

corresponds to the presence of the pseudo-tensor density t_μ^ν in the equation in this form (16).

This general result proves to be of great importance for relativistic thermodynamics by permitting—even in the case of isolated systems—an increase in the energy directly associated with matter and radiation at the expense of the potential energy that we assign to the gravitational field. For example, if we consider a system composed of a perfect fluid having the proper macroscopic density of energy ρ_{00} and proper pressure p_0 as measured by a local observer at the point of interest, and having no flow of heat, it is known that we can write

$$T^{\mu\nu} = (\rho_{00} + p_0)\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} - g^{\mu\nu}p_0 \tag{17}$$

as an expression for the energy-momentum tensor. And if we substitute this expression into the above equations of relativistic mechanics, we can obtain for any infinitesimal element of the fluid of proper volume δV_0 the relation

$$\frac{d}{dt_0}(\rho_{00}\delta V_0) + p_0\frac{d}{dt_0}(\delta V_0) = 0. \tag{18}$$

From one point of view there is nothing surprising about this result since it merely states that a local observer who examines the behavior of an element of the fluid small enough so that the gravitational *curvature* of space-time can be neglected will find the rate of change in energy content related in the expected way to the work done against the external pressure. From another point of view, however, as this same equation can be applied to each one of all the elements into which the total fluid of the system can be divided, the result may seem somewhat surprising, since it leads to the possibility of systems in which the proper energy of every element of the fluid may be simultaneously decreasing or increasing, according as the system is expanding or contracting. Moreover, since it is this proper energy immediately associated with matter and radiation which determines the possibilities for entropy increase, we shall later find in relativistic thermodynamics an escape from certain restrictions imposed in the classical thermodynamics by the usual form of the principle of the conservation of energy.

Just as in the previous case of special relativity, we note that the extension of thermodynamics to general relativity involves, so far as the first law is concerned, no new hypothetical material beyond that already contained in relativistic mechanics. And we may now turn to the relativistic analogue of the second law of thermodynamics.

b. *The Analogue of the Second Law in General Relativity.* To guide us in obtaining a suitable postulate to serve as the relativistic second law of thermodynamics, we must make use of the two fundamental ideas of general relativity which are expressed by the principles of covariance and equivalence. In accordance with the principle of covariance, our postulate must be expressed in covariant form the same for all coordinate systems, to avoid the danger of being influenced in its selection by a spurious simplicity when referred to some particular system of coordinates. And in accordance with the principle of equivalence, our postulate must reduce to the thermodynamic requirements of special relativity, when applied to an infinitesimal element of fluid, using natural coordinates for the point of interest.

These two principles have been sufficient to lead with considerable confidence to the expression*

$$\frac{\partial}{\partial x^\mu}\left(\phi_0 \frac{dx^\mu}{ds}(-g)^{1/2}\right)\delta x_1\,\delta x_2\,\delta x_3\,\delta x_4 \geqq \frac{\delta Q_0}{T_0} \tag{19}$$

as the appropriate postulate to take as the relativistic analogue of the ordinary second law of thermodynamics. The quantity ϕ_0 in this expression is the proper entropy density of the thermodynamic fluid under consideration as measured at the point of interest by a local observer; the quantities dx^μ/ds are the components of the macroscopic *velocity* of the fluid at that point; and the other quantities on the left-hand side of the expression have their usual significance. The significance of the right-hand side of the expression is more difficult to grasp, and will be specially treated in a forthcoming article by Robertson and myself.† The quantity δQ_0 may be taken as the heat, measured by a local observer at rest in the fluid at the point of interest, which

* Tolman, Proceedings of the National Academy, vol. 14 (1928), pp. 268, 701; Physical Review, vol. 35 (1930), p. 896.

† Tolman and Robertson, submitted to the Physical Review.

flows into an element of the fluid having the instantaneous proper volume δV_0 during the proper time δt_0, where these quantities are so chosen as to make

$$\delta V_0 \, \delta t_0 = \sqrt{-g} \, \delta x_1 \, \delta x_2 \, \delta x_3 \, \delta x_4, \tag{20}$$

and the quantity T_0 is taken as the temperature ascribed to this heat by the local observer.

The two signs of equality ($=$) and inequality ($>$) in the expression refer respectively to the two cases of reversible and irreversible processes, and in applying the principle to irreversible processes we are to regard an increment in coordinate time δx_4 as positive when taken in the direction to correspond to a positive increment in proper time δt_0 as measured in the ordinary manner by a local observer.

To show the agreement of this postulated expression for the relativistic second law with the principle of covariance, we have merely to note that it is a tensor equation of rank zero—both sides being scalar invariants—and hence is true in all coordinate systems if true in one. To show its agreement with the principle of equivalence we must see what it reduces to in natural coordinates for the point of interest. Introducing such coordinates x, y, z, t, *and making use of the transformation equations for entropy, heat, and temperature provided by the special theory of relativity*, we find, however, that our principle then reduces to

$$\left[\operatorname{div} (\phi \bar{u}) + \frac{\partial \phi}{\partial t} \right] \delta x \, \delta y \, \delta z \, \delta t \geqq \frac{\delta Q}{T}, \tag{21}$$

where ϕ, $\bar{u}$, δQ and T are now the quantities referred to our present coordinate system which we ordinarily designate as entropy density, velocity, heat absorbed, and temperature. And we see that this result does relate the change in the entropy of the element of fluid, instantaneously contained in the coordinate range $\delta x \, \delta y \, \delta z$, to the absorbed heat and temperature in the way

$$\frac{d}{dt}(\phi \, \delta x \, \delta y \, \delta z) \delta t \geqq \frac{\delta Q}{T}, \tag{22}$$

which is required by the second law of thermodynamics in special relativity.

At the present stage of observational knowledge, our belief in the validity of the proposed postulate is primarily based on

this agreement with the two principles of covariance and equivalence. In addition, however, it may be emphasized that the principle has been chosen so as to be simply the immediate covariant re-expression of the special relativity form of the second law; and past experience has shown, notably for example in the cases of the fundamental formulas for space-time interval and geodesic trajectory, that these simplest possible covariant generalizations, when feasible, are likely to be correct. Furthermore, it may be remarked that the conclusions which have so far been drawn from this extension of thermodynamics to general relativity appear—at least after due reflection—to be reasonable and illuminating.

It must be emphasized, nevertheless, that these qualities are not sufficient to prove the validity of the postulate, since other covariant expressions might be found which would also reduce to the special relativity law in natural coordinates. Hence the postulate must be regarded as a real generalization with a range of validity to be finally determined only by the correspondence between observation and prediction.

5. *Consequences of Relativistic Thermodynamics.* To complete our discussion, we must now consider the possible consequences of relativistic thermodynamics. The technical modifications in thermodynamic theory needed to secure its extension to relativity may have seemed too trivial and obvious to warrant the expectation that these consequences could be very novel or interesting. Nevertheless, the actual effect of replacing classical ideas as to the nature of space and time by relativistic ideas is so fundamental as to lead to important differences between the results of classical and relativistic thermodynamics. Three examples may now be given to illustrate both the essential novelty and the inherent rationality of the consequences of relativistic thermodynamics.

a. *Temperature Gradient in Gravitational Field.* In the classical thermodynamics we have become accustomed to the conclusion that a system which is in thermal equilibrium will necessarily have uniform temperature throughout. As a result of relativistic thermodynamics, however, it is found that this conclusion must be modified in the presence of appreciable gravitational fields. Thus if we consider a spherical distribution of material held to-

gether in static equilibrium by its own gravitational forces, corresponding to the line element

$$ds^2 = -e^{\mu}(dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2) + e^{\nu} dt^2, \tag{23}$$

where μ and ν are functions of r alone, it can be shown* that the condition for thermal equilibrium is given by

$$\frac{d \log T_0}{dr} = -\frac{1}{2}\frac{d\nu}{dr}, \tag{24}$$

where T_0, the proper temperature as measured by a local observer, decreases as we go outward instead of remaining constant. And for the more general static case of a system, not necessarily having spherical symmetry, but corresponding to the more general line element

$$ds^2 = g_{ij}\, dx^i\, dx^j + g_{44} dt^2, \qquad (i, j = 1, 2, 3), \tag{25}$$

where g_{ij} and g_{44} are independent of t, it has been shown by Ehrenfest and myself† that the condition for thermal equilibrium would be given by a constant value throughout the system for the product

$$T_0 (g_{44})^{1/2} = \text{const.} \tag{26}$$

rather than for the proper temperature T_0 itself.

From an observational point of view, this new conclusion may not be very important, since the change in temperature with position would only be

$$\frac{d \log T}{dr} = 10^{-17}\ \text{cm}^{-1} \tag{27}$$

in a gravitational field having the intensity of that at the earth's surface. From a theoretical point of view it is of interest, however, in leading to a modification of one of the most cherished results of the classical thermodynamics.

It should be emphasized, nevertheless, that this modification is entirely reasonable from the point of view of relativity. In accordance with that theory all forms of energy have mass and weight, and it is hence indeed not surprising that a temperature

* Tolman, Physical Review, vol. 35 (1930), p. 904.

† Tolman and Ehrenfest, Physical Review, vol. 36 (1930), p. 1791.

gradient is necessary to prevent the flow of heat from regions of higher to those of lower gravitational potential when thermal equilibrium has been reached. Furthermore, since the pressure of black body radiation in equilibrium with a thermal system would evidently have to increase as we go to lower gravitational levels, in order to support the weight of radiation above, we can also see from purely mechanical considerations that temperature must increase with depth.

b. *Possibility for Reversible Processes at a Finite Rate.* To turn to a second example, we have long been accustomed to believe as a result of the classical theory that *reversible* thermodynamic processes taking place at a *finite rate* could never occur in nature, since a finite rate of change would make it impossible to achieve that maximum efficiency which would permit a restoration both of the system and its surroundings to their original states.

To illustrate this we may consider a fluid contained in a cylinder with non-conducting walls and provided with a piston as shown in Fig. 1. The *reversible* expansion of this fluid with the piston moving out at a *finite rate* would evidently be impossible, in the first place since there would be friction between the piston and the walls, and in the second place since the fluid in flowing in behind the moving piston would not be able to maintain as high a pressure and hence do as much external work as at an infinitesimal rate of expansion. It has hence been concluded in the past that the reversible expansion of a fluid at a finite rate would under all circumstances be impossible.

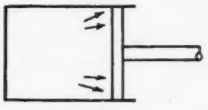

Fig. 1
Classical Cylinder

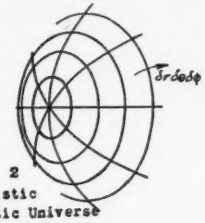

Fig. 2
Relativistic
Non-static Universe

In relativistic thermodynamics, nevertheless, this situation is altered in an important manner by new possibilities for changes to take place in the proper volume of an element of fluid, be-

cause of changes in the gravitational potentials which were neglected in the older theory. To illustrate this, Fig. 2 attempts to give a symbolic two-dimensional representation of the space-like coordinates, corresponding to the line element

$$(28)\quad ds^2 = -\frac{e^{g(t)}}{\left[1+\frac{r^2}{4R^2}\right]^2}(dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2) + dt^2$$

which characterizes the homogeneous distribution of fluid that we usually call a non-static model of the universe. These coordinates are so selected that any element of fluid lying in a given coordinate range $\delta r\ \delta\theta\ \delta\phi$ will remain permanently therein. The proper volume of such an element of the fluid as measured by a local observer will then be

$$(29)\quad \delta V_0 = \frac{e^{(3/2)\,g(t)}}{\left[1+\frac{r^2}{4R^2}\right]^3} r^2 \sin\theta\ \delta r\ \delta\theta\ \delta\phi,$$

and this will change with the time t at a finite rate on account of the occurrence of the function $g(t)$ in the gravitational potentials and hence also in the expression for proper volume.

These changes in proper volume take place, however, in the first place quite obviously without any friction of moving parts, and in the second place with a perfect balance between internal and external pressures owing to the uniformity of conditions throughout the fluid. Thus the two sources of irreversibility in the previous classical illustration are completely eliminated. Furthermore, there can be no irreversible heat flow in the model under consideration, again owing to the uniform conditions throughout the fluid, and the possibility of irreversible processes within each element of the fluid can be eliminated by the choice of a sufficiently simple substance. As a consequence it has proved possible to construct conceptual models of this kind* which can expand or contract at a finite rate, and nevertheless satisfy the relativistic condition for reversibility by maintaining the sign of equality instead of inequality in the simple expression

* Tolman, Physical Review, vol. 37 (1931), p. 1639; ibid., vol. 38 (1931), p. 707.

$$\frac{d}{dt}(\phi_0 \delta V_0) \geqq 0, \tag{30}$$

which results from the application of the general form of the relativistic second law (19) to the present case.

This possibility for reversible thermodynamic processes taking place at a finite rate seems very strange from the point of view of the classical thermodynamics. Nevertheless, from the relativistic point of view the result seems both rational and, indeed, perhaps inevitable, when we recall that the principles of *relativistic mechanics* alone, without bothering about thermodynamics, have been found sufficient to show* the possibility of constructing models of the kind in question, which—when filled with such simple fluids as a perfect monatomic gas or black-body radiation—would expand at a finite rate to an upper limit and then precisely retrace their steps to their original density, pressure, and temperature.

c. *Possibility for Irreversible Processes without Reaching a Final State of Maximum Entropy.* As a final example of the differences between classical and relativistic thermodynamics, let us now turn to the classical conclusion that the ultimate result of irreversible processes taking place in a system having no interaction with its surroundings would necessarily be a state of maximum entropy where further change would be impossible.

To illustrate the classical reasons for belief in this principle, let us consider a simple homogeneous system consisting of a gas of uniform pressure, temperature, and composition throughout. The state of such a system can be specified by its energy E, volume v, and the number of mols N_1, N_2, etc., of the different chemical substances which it contains, and in accordance with a fundamental expression, provided by the work of Gibbs† himself, its entropy can then be determined with the help of the general equation

$$dS = \frac{dE}{T} + \frac{p}{T}dv + \left(\frac{\partial S}{\partial N_1}\right)dN_1 + \left(\frac{\partial S}{\partial N_2}\right)dN_2 + \cdots. \tag{31}$$

* Einstein, Berlin Berichte (1931), p. 235; Tolman, Physical Review, vol. 38 (1931), p. 1758.

† Allowing for the difference in notation and form, equation (31) is equivalent to the fundamental equation (12) given in the *Collected Works of Gibbs*, Longmans, Green and Company, 1928, vol. I, p. 63.

In applying this expression, however, to the special case of a system having no interaction with its surroundings, it was necessary in the classical thermodynamics to take the energy change dE equal to zero to agree with the classical principle of energy conservation. It was also necessary, for a system having no interaction with its surroundings, to take either the volume-change dv equal to zero or the pressure p itself equal to zero to prevent the performance of work on the surroundings. In both cases the only chance for entropy increase and hence for change would then lie in the readjustment of chemical composition. At constant energy and volume, moreover, this would have to cease when the entropy reached the maximum possible value compatible with the fixed value of energy and volume; while at constant energy and zero pressure the gas would be infinitely dilute and a final state of maximum entropy would be reached when all the molecules of gas had disassociated into their atoms.

In the relativistic treatment of this problem, nevertheless, the application of relativistic mechanics alone is sufficient to show a very different state of affairs. In the relativistic treatment, the analogue of the preceding homogeneous isolated system will evidently be a non-static model of the universe, since the line element of these models corresponds to a completely homogeneous distribution of fluid with no interaction with anything outside the system itself. By the application of relativistic mechanics, however, it has been shown by the work of Ward and myself,* that there is a great class of such models, obtained by taking the cosmological constant Λ equal to zero or less, which could apparently undergo a continued succession of expansions and contractions without ever coming to rest. To be sure, the equations that we now have available for our highly idealized models are only sufficient to describe the expansion of the models to their upper limit and return, and not sufficient to describe their passage through the exceptional point at their lower limit of volume. Nevertheless, on physical grounds we must be inclined to assume that contraction to the lower limit of volume would be followed by renewed expansion.

It is to be specially noted now, however, that our conclusion

* Tolman, Physical Review, vol. 30 (1932), p. 320; Tolman and Ward, ibid., vol. 39 (1932), p. 835.

as to the possibility of such behavior is based on mechanical equations alone and is quite independent of assumptions as to the reversibility or irreversibility of any thermodynamic processes that may take place in the fluid, although in the case of irreversible processes it can be shown that successive expansions have a tendency to go to larger and larger volumes before returning. The possibility thus provided for an unending succession of irreversible expansions and contractions seems very strange from the point of view of classical thermodynamics, but can nevertheless be shown to be reasonable from the point of view of relativistic thermodynamics.

The application of the relativistic second law to these non-static models leads to the general result

$$\frac{d}{dt_0}(\phi_0 \delta V_0) \geqq 0, \tag{32}$$

which states that the proper entropy of each element of the fluid as measured by a local observer can only remain constant or increase with local time. With a fluid of composition simple enough to eliminate the possibility for changes other than in density, we obtain the conditions necessary for constant entropy, and are led to the reversible expansions and contractions at a finite rate previously considered. With a slightly more complicated fluid, however, such as a gas which tends to disassociate as the density is lowered, we can obtain the irreversible increases in entropy in which we are now interested, since the composition of the fluid will then lag behind the changes in volume, and the processes of disassociation and recombination will always take place in the direction of an equilibrium which has not been attained.

From a classical point of view, such a continuous irreversible increase in entropy in an isolated system which undergoes a never-ending succession of expansions and recontractions to an earlier volume might seem impossible, since the energy of the isolated system would have to remain constant, and with a given value of energy and volume there would be a definite, maximum possible value of entropy.

From the relativistic point of view, nevertheless, it is easy to see that this continued increase in entropy could be made possible by the failure in relativistic mechanics of the ordinary prin-

ciple of energy conservation. Indeed the application of relativistic mechanics to the models in question leads to the result

$$\frac{d}{dt_0}(\rho_{00}\delta V_0) + p_0 \frac{d}{dt_0}(\delta V_0) = 0, \tag{33}$$

which shows that the proper energy $(\rho_{00}\delta V_0)$ of each element of fluid in the model will decrease as the model expands and increase as it contracts, and since the general effect of irreversibility would be to give higher pressures during compression than during the preceding expansion, we are thus provided with the needed mechanism for a gradual increase in proper energy and hence also in the proper entropy of the elements of fluid.

The difference between the classical and relativistic points of view can also be illustrated in a somewhat different manner, if we return to the classical cylinder and relativistic model of the universe given by Figures 1 and 2, and consider a continued succession of expansions and contractions, in both cases providing for irreversibility by again introducing for example a gas which tends to disassociate when the pressure is lowered.

The classical man who is engaged in moving the piston of his cylinder successively in and out, reports that the entropy inside the cylinder is all the time getting greater and greater, because the gas cannot disassociate fast enough on the way out nor recombine fast enough on the way in to maintain equilibrium. He also reports, however, that the energy inside his cylinder—which is now no longer an isolated system—is also getting greater and greater, since the pressure has a tendency to be too high to correspond to equilibrium on compression and too low on expansion, so that he does more work on the way in than he gets back on the way out. He then finally remarks that this continued increase in energy indeed provides in the case of this non-isolated system an opportunity for the entropy to go on increasing forever, but that he personally is getting very tired of the silly experiment, and is going to have to stop:—not now because the irreversible processes will ever lead to an unsurpassable maximum of entropy but because he actually won't have the necessary energy to push in the piston one time more.

How different the remarks of the relativistic man as he sits calmly by and—in his mind's eye—watches his conceptual model expand and contract. "Yes," he says, "I notice that the

proper entropy of each element of fluid is all the time increasing owing to successive processes of disassociation and recombination under conditions that do not correspond to equilibrium." "I do not worry, however, since I also notice that the proper energy of each element of fluid is also increased after a succession of expansions and contractions sufficiently to allow for this increase in entropy, and I know that the principles of relativistic mechanics not only permit such increases in proper energy, but indicate—without reference to reversibility or irreversibility—the continued succession of expansions and contractions that I observe."

This completes the discussion of examples illustrating the kind of conclusions that we may expect from the extension of thermodynamics to general relativity.

6. *Conclusion.* Much remains to be done in the further application and development of relativistic thermodynamics. The application to systems in which heat flow is taking place from one portion of the fluid to another has not yet been undertaken and might lead to interesting results. It could be of special importance when we have made more progress in the treatment of non-homogeneous models of the universe which may well be necessary for a better understanding of the actual universe. The development of relativistic thermodynamics to include an appropriate treatment of fluctuations might also be attempted with the help of the statistical methods introduced by Gibbs himself. The effect of fluctuations on the behavior of cosmological models might be specially important at certain stages of their history.

In trying to estimate the significance of the applications of relativistic thermodynamics that have already been made, we must not misjudge the nature of the two applications to cosmological models that were described in the foregoing. It should be emphasized that the homogeneous cosmological models which we now consider are not only very highly simplified and idealized, but at best are constructed to agree throughout their entire extent with that small sample of the actual universe which lies within the range of some 10^8 light years. Furthermore, it must be remembered that among the different possible kinds of homogeneous cosmological model, there is a class which would

expand never to return as well as the class that could undergo an unending succession of expansions and contractions; and we do not now have sufficient data so that we could assign the actual universe to either class. Hence we must be very careful in extrapolating to the actual universe any conclusions that we may draw as to the behavior of our conceptual models.

The dangers of long-range scientific extrapolation were specially emphasized to you by Professor Bridgman in his beautiful Gibbs memorial lecture of last year, and these dangers, as we see, are specially present in cosmological speculations, based on the observation of a small fragment of the total universe for an inappreciable time span. Nevertheless, it certainly seems significant that conceptual models of the universe can be constructed, which are permitted by the principles of relativistic thermodynamics to exhibit behavior in serious conflict with the classical conclusions, that reversible processes could never take place at a finite rate, and that the end result of irreversible processes would necessarily be a stationary condition of maximum entropy. Hence at the very least, it would now seem desirable to extrapolate to a cautious position, in which we no longer dogmatically assert that the principles of thermodynamics necessarily require a universe created at a finite time in the past and fated for stagnation and death in the future. Indeed, the chief duty and glory of theoretical science is certainly not merely to describe in complicated language those facts that are already known but to extrapolate—as cautiously and wisely as may be—into regions yet unexplored but pregnant with human interest.

California Institute of Technology

NON-ANALYTIC FUNCTIONS OF A COMPLEX VARIABLE*

BY E. R. HEDRICK

1. *Introduction.* The outlines of the theory of non-analytic functions of a complex variable, called also polygenic functions, have been stated in recent years in a number of articles.† Indeed, from the very first of the modern study of the theory of functions, going back at least as far as the famous inaugural dissertation of Riemann, the beginnings of the subject have been mentioned essentially, if for no other purpose than to state the conditions under which a function of a complex variable is analytic, and to delimit the field of functions to be studied.

In the present address, such preliminary details will be mentioned only briefly, with references; but enough of them must be stated to develop a notation, and to give the proper setting. More detailed attention will be given to those developments which have taken place during the last decade, and to some hitherto unpublished facts. A brief review of some of the historical background will serve both its obvious purpose, and also that of introducing the necessary preliminary details and notations.

2. *Historical Background.* As was stated above, every careful presentation of the classical theory of functions of a complex variable did include in a measure the elementary ideas for the general case of *any* function of a complex variable. A function

$$w = f(z) = \phi(x, y) + i\psi(x, y), \tag{1}$$

where $w=u+iv$ and $z=x+iy$, is said to be defined for a given region (or set of values) of z if w is determined whenever z is assigned a value in that region (or set). The equation (1) is then equivalent, of course, to the two real simultaneous equations

$$u = \phi(x, y), \quad v = \psi(x, y), \tag{2}$$

which themselves express a transformation of the xy plane onto

* An address delivered at New Orleans, December 31, 1931, as the retiring presidential address, before the American Mathematical Society.

† See the list of recent articles at the end of this paper.

the uv plane. In general, we shall assume that ϕ and ψ are continuous, and that they possess continuous first partial derivatives, unless the contrary is stated.

If z and w have corresponding increments Δz and Δw, the limit of their ratio, $\Delta w/\Delta z$, is the derivative dw/dz whenever it has a unique value for all methods of approach of Δz to zero, for a fixed value of z; in general, that limit will depend upon the slope m of the curve along which Δz approaches zero, and we shall call it *the directional derivative*. This limit is written in a variety of forms in standard works; we shall write

$$\gamma = \alpha + i\beta = \frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\Delta u + i\Delta v}{\Delta x + i\Delta y} = \frac{u_x + iv_x + m(u_y + iv_y)}{1 + im}, \tag{3}$$

where u_x, u_y, v_x, v_y denote the partial derivatives in the usual manner, and where $m = dy/dx$ is the slope of the curve of approach. In order that (3) give a unique result independent of m, the usual necessary and sufficient conditions are the well known Cauchy-Riemann equations

$$u_x = v_y, \quad u_y = -v_x. \tag{4}$$

The equation (3) may be written in other forms. Riemann, in his inaugural dissertation, wrote it essentially in the following form. Let θ denote the angle between the x axis and the curve of approach, so that $\tan\theta = m$; then, by means of the usual relation $e^{i\theta} = \cos\theta + i\sin\theta$, it is easy to reduce (3) to the form

$$\gamma = \frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \mathcal{D}[f(z)] + \mathcal{P}[f(z)]e^{-2i\theta}, \tag{5}$$

where

$$\begin{aligned} \mathcal{D}[f(z)] &= \tfrac{1}{2}[u_x + v_y + i(v_x - u_y)], \\ \mathcal{P}[f(z)] &= \tfrac{1}{2}[u_x - v_y + i(v_x + u_y)]. \end{aligned} \tag{6}$$

These expressions will occur frequently in what follows. The expression $\mathcal{D}(w)$, which is the part of (5) that is independent of θ, is precisely the value of the derivative dw/dz whenever the derivative has a unique value at the point z, that is, whenever the

result in (5) is independent of θ. The expression for $\mathcal{P}(w)$ vanishes when and only when the equations (4) hold. It is pointed out that some of the results recently announced can be seen as immediate consequences of (5).

The preceding classical statements evidently do not presume *in advance* that w is an analytic function of z, nor that the equations (4) hold. This is true of many other results of the classical theory. In particular, the well known Morera Theorem,* the Weierstrass Mean-Value Theorem,† and many others, do not presume the analytic character of the functions in their hypotheses. In such theorems as that of Morera, in fact, the very point of the theorem would be lost if there were such an original hypothesis.

Specific attempts to discuss as such functions that are not analytic have been made from time to time. One such that I would mention is a paper by Picard,‡ in which the attention is directed toward generalizing the Cauchy-Riemann equations (4).

The expression $\mathcal{P}(w)$ defined in (6) is equal, except for a constant factor (see §9), to a concept considered by D. Pompeiu, and later by Hayashi.§ Pompeiu arrived at this concept, which he called the *areal derivative*, by considering the limit

$$\lim_{C=0} \frac{\int_C f(z)dz}{\frac{1}{2}\int_C xdy - ydx} = \lim_{C=0} \frac{\int_C f(z)dz}{A}, \tag{7}$$

where C denotes a contour surrounding the point P, A the area of the region bounded by C, and where the limit is taken as C shrinks toward P in such a way that C eventually lies in an arbitrarily small circle about P. Some of the results of these papers are stated below (§9).

* Morera, Rendiconti, Istituto Lombardo, (2), vol. 19 (1886); and texts on the classical theory.

† See, for example, Osgood, *Lehrbuch der Funktionentheorie*, 1923, vol. 1, p. 212; Goursat, *Mathematical Analysis* (English ed.), vol. 2, Part 1, p. 65.

‡ Journal de Mathématiques, (4), vol. 8 (1892), p. 217.

§ D. Pompeiu, Rendiconti di Palermo, vol. 33 (1912), pp. 108–113, and vol. 35 (1913), p. 277.

The other papers cited at the end of this address, and still others, have appeared during the past decade. They form together a body of knowledge concerning this field of which the remainder of this address gives a summary.

3. *Principal Directions. Characteristic Lines.* If we introduce the familiar notation

$$E = u_x^2 + v_x^2, \quad G = u_y^2 + v_y^2, \quad F = u_x u_y + v_x v_y, \tag{8}$$

it follows readily* that

$$r = \left|\frac{dw}{dz}\right|^2 = \left|\lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z}\right|^2 = \frac{du^2 + dv^2}{dx^2 + dy^2} = \frac{E + 2mF + Gm^2}{1 + m^2}, \tag{9}$$

where m denotes, as before, the slope dy/dx of the curve of approach. The maximum and minimum values of r are found to exist for values of m given by the equation

$$F + (G - E)m - Fm^2 = 0, \tag{10}$$

and these values of r are themselves solutions of the quadratic equation

$$\rho^2 - (E + G)\rho + J^2 = 0, \tag{11}$$

where

$$J = (EG - F^2)^{1/2} = u_x v_y - u_y v_x.$$

Since the roots of (10) are negative reciprocals, it follows that the directions which give a maximum and a minimum *stretching ratio* $R = r^{1/2}$, are *at right angles*. These directions we called the *principal directions*. They were discovered first by Tissot† in his study of the mapping problem. The two families of curves that are tangent to these principal directions at every point we called the *characteristic curves* for the given function.

It is shown‡ that *the only orthogonal system of curves in the z plane which corresponds by the transformation $w = f(z)$ to an orthogonal system in the w plane, is the system of characteristic lines.*

If we lay off, at a fixed point z_0, in the direction m, the recipro-

* See Hedrick, Ingold, and Westfall, loc. cit.

† Tissot, *Sur les cartes géographiques*, Comptes Rendus, 1849.

‡ Hedrick, Ingold, and Westfall, loc. cit. See also the geometric proof given by Darboux, *Leçons sur la Théorie des Surfaces*, vol. 3, p. 49.

cal stretching factor $P=1/R$, the curve obtained* is the ellipse

$$E\bar{x}^2 + 2F\bar{x}\bar{y} + G\bar{y}^2 = 1, \tag{12}$$

where $\bar{x}$, $\bar{y}$ are the coordinates measured from the point z_0 as origin. This ellipse may be called the *Tissot indicatrix*. Similarly, if (2) can be solved for x and y in terms of u and v, the factor R can be laid off about the point w_0 in the w plane to obtain the corresponding ellipse

$$\mathcal{E}\bar{u}^2 + 2\mathcal{F}\bar{u}\bar{v} + \mathcal{G}\bar{v}^2 = 1 \tag{13}$$

in the uv plane, where $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$ are the fundamental quantities similar to E, F, G when u and v are the independent variables.

It is clear that these concepts all degenerate, the ellipses becoming circles and the principal directions becoming indeterminate, if

$$E = G, \quad F = 0. \tag{14}$$

These equations are satisfied, of course, if the Cauchy-Riemann equations (4) are satisfied, that is, *if $f(z)$ is an analytic function of z*, or also if $f(z)$ is an analytic function of $\bar{z}=x-iy$, in which case the signs in (4) are changed, but (14) still hold.

4. *Pointwise Analytic Functions.* It is evident that the derivative dw/dz may exist at isolated points in the sense that the limit (3) may be independent of m at isolated points. These points are the solutions of (4) regarded as simultaneous equations for x and y. We say that *$f(z)$ is analytic at a point z_0* if (4) hold at that point.

Since we have assumed that u and v, and their first derivatives, are continuous, it follows that *the points at which $f(z)$ is analytic form a closed set*, except for points at which $f(z)$ is not defined or does not satisfy these hypotheses. At all points at which $f(z)$ is analytic, the ellipses defined in §3 become circles and the principal directions become indeterminate.

5. *The Beltrami Equations. Classification of Functions.* The Cauchy-Riemann equations (4) have been generalized by Beltrami† to the case of any surface. These equations are

* Hedrick, National Academy of Sciences, loc. cit.

† Beltrami, *Delle variabili complesse sopra una superficie qualunque*, Annali di Matematica, (2), vol. 1; see also Picard, *Traité d'Analyse*, vol. 2, p. 8.

$$Jv_x = Fu_x - Eu_y,$$
$$Jv_y = Gu_x - Fu_y; \tag{15}$$

and u and v satisfy the generalized Laplace equation

$$\frac{\partial}{\partial x}\left(\frac{FU_y - GU_x}{J}\right) + \frac{\partial}{\partial y}\left(\frac{FU_x - EU_y}{J}\right) = 0. \tag{16}$$

These equations, which evidently hold for the plane also, are here written with the symbol $J=(EG-F^2)^{1/2}$, instead of the symbol H employed in the theory of surfaces.

As in the case of a surface, so here also, those functions for which the ratios $E:F:G$ are the same form a *class*. Every function $f(z)$ belongs to some class; no function belongs to two different classes; there exist functions belonging to every class; and if two functions belong to the same class, the one is an analytic function of the other.

It is easy to prove* that two functions $f=u+iv$ and $F=U+iV$ do belong to the same class, so that one is an analytic function of the other, if and only if the jacobian of F and f with respect to x and y vanishes:

$$\begin{vmatrix} \dfrac{\partial(U+iV)}{\partial x} & \dfrac{\partial(U+iV)}{\partial y} \\ \dfrac{\partial(u+iv)}{\partial x} & \dfrac{\partial(u+iv)}{\partial y} \end{vmatrix} = 0. \tag{17}$$

For functions of the same class, a function theory entirely analogous to that of ordinary analytic functions exists.

6. *The Kasner Circle.* Let us now return to the formula for the directional derivative, which we write in the form

$$\gamma = \alpha + i\beta = \left.\frac{dw}{dz}\right|_\theta = \mathcal{D} + \mathcal{P}e^{-2i\theta}, \tag{18}$$

where $\mathcal{D}$ and $\mathcal{P}$ are functions of z alone, independent of θ, and are defined by the equations (6). It is immediately evident from (18) that the values of the directional derivative γ, for a fixed value of z, depend only on θ, and that *these values of γ all lie on a*

* See Nicolesco, Thesis, loc. cit.

circle in the γ plane, whose center is the point $\mathcal{D}(w)$ and whose radius is $|\mathcal{P}(w)|$. In fact, (18) can be written in the form

$$\gamma - \mathcal{D} = \mathcal{P}e^{-2i\theta}, \tag{19}$$

whence

$$|\gamma - \mathcal{D}| = |\mathcal{P}|, \tag{20}$$

which is the equation of the circle described above. Here $\gamma - \mathcal{D}$ is a vector counted from the point $\mathcal{D}$ as origin, which is a fixed point if z is fixed. If we insert the detailed values given by (6), and the value $\alpha + i\beta$ for γ, the equation (20) becomes

$$\left[\alpha - \left(\frac{u_x + v_y}{2}\right)\right]^2 + \left[\beta - \left(\frac{v_x - u_y}{2}\right)\right]^2 = \left(\frac{u_x - v_y}{2}\right)^2 + \left(\frac{v_x + u_y}{2}\right)^2. \tag{21}$$

In this form, the equation was first published by Kasner.* I have proposed that the circle be called the Kasner circle. His original derivation of the equation (21) is not so immediate, however, as is the process given above; all that is necessary is to transpose the term $\mathcal{D}$ in the Riemann equation (5) and then eliminate θ by taking the absolute values on both sides.

It is obvious also from (19) that, *as θ varies in the z plane, the point γ moves around the circle at twice the rate at which θ changes, and in the opposite sense.*

On account of the resemblance just described to an ordinary clock-motion, Kasner has proposed that the distribution of the values of the directional derivative on this circle be called a *clock*. As z varies, there will be one such clock for every value of z in the z plane; Kasner has called the whole family a *congruence of clocks*. He has pointed out also† that the mean value of γ on the circle is

$$\frac{1}{\pi}\int_0^\pi \gamma d\theta = \frac{1}{\pi}\int_0^\pi (\alpha + i\beta)d\theta = \frac{1}{\pi}\int_0^\pi (\mathcal{D} + \mathcal{P}e^{-2i\theta})d\theta = \mathcal{D}; \tag{22}$$

* Kasner, Science, vol. 66 (1927), pp. 581–582.

† Kasner, National Academy of Sciences, loc. cit., p. 81.

and he calls $\mathcal{D}(w)$ the *mean derivative* of w. The reason for this result is simply that $\mathcal{D}$ *and* $\mathcal{P}$ *are constant during this integration*, since only θ varies; and since

$$\int_0^\pi \mathcal{P}e^{-2i\theta}d\theta = \mathcal{P}\int_0^\pi e^{-2i\theta}d\theta = 0, \tag{23}$$

the term involving $\mathcal{P}$ in (22) falls out, and (22) becomes an identity.

Other developments regarding the congruence of clocks as a picture of the directional derivative are given by Kasner (loc. cit.) and by Kasner and Hofmann.*

7. *Riemann Surfaces. The Increment Ratio.* It is well known that any function $w=f(z)$, that is, any transformation of the form (2), can be represented by a Riemann surface. The equations (2) can be solved for x and y under the hypotheses made above, near any point (x_0, y_0) at which the jacobian $J=u_xv_y-u_yv_x$ does not vanish. If there are several solutions near (u_0, v_0), they can be represented geometrically by points of a Riemann surface of several sheets over the uv plane, as in the usual theory. These sheets must be connected, if at all, along the *branch curves* K in the uv plane, which are curves that correspond to the critical curve

$$J(x, y) = 0 \tag{24}$$

in the xy plane. The ordinary *branch points* of the classical theory of functions are degenerate cases of these branch curves, since the jacobian reduces to a sum of squares if the Cauchy-Riemann equations (4) hold. *Any curve in the xy plane that passes through a point of the curve* (24) *corresponds to a curve in the uv plane that is tangent to the corresponding branch curve* K, *unless the corresponding point* (u, v) *is a singular point of* K.†

On account of these properties, I have proposed to call the branch curves K the *edge of regression* on the Riemann surface, since they play a role very similar to that of the edge of regression of a developable surface. Indeed, *if any family of curves*

$$x = p(t, \alpha), \quad y = q(t, \alpha),$$

* This Bulletin, vol. 34 (1928), pp. 495–503.

† Hedrick, Bulletin of the Calcutta Society, loc. cit., p. 114.

crosses the critical curve C at a variable point that is a function of α, the edge of regression K is the envelope of the corresponding family in the uv plane.

I have stated some examples in my Calcutta paper (loc. cit.).

The usual derivative ratio

$$\zeta = \xi + i\eta = \frac{\Delta w}{\Delta z} = \frac{\Delta u + i\Delta v}{\Delta x + i\Delta y} \tag{25}$$

is a single-valued function of $z=z_0+\Delta z$ for a fixed point z_0, except at the point $z=z_0$, at which it is not defined. I have shown in my Calcutta paper that the jacobian of ζ

$$J = \xi_x\eta_y - \xi_y\eta_x$$

vanishes at precisely the points of the Kasner circle. Hence the Kasner circle is precisely the edge of regression for the function ζ, which is in general represented on a two-leaved Riemann surface whose leaves join along the Kasner circle. As the point z revolves once around the point z_0, the point ζ revolves twice about the Kasner circle.

This fact makes it very evident why the directional derivative γ revolves on the Kasner circle twice as fast as the angle θ revolves in the z plane; the directional derivative is indeed precisely the edge of regression, geometrically, and is the missing definition of what corresponds in (25) to the value $z=z_0$.

8. *Other Geometric Properties.* Several particular cases, and particular values of the directional derivative γ on the circle, are given by Kasner.* He uses the fact that the point $A(v_y, v_x)$ lies on the circle. Other important special points on this circle are $E(u_x, v_x)$, $D(v_y, -u_y)$, $B(u_x, -u_y)$. In my Calcutta paper, I have used these four points A, E, D, B, which evidently lie at the vertices of a rectangle inscribed in the circle. From the origin O in the plane of γ, the distances OA and OE are

$$OA^2 = v_x^2 + v_y^2 = G, \quad OE^2 = u_x^2 + u_y^2 = E,$$

which represent the two fundamental quantities E and G. It is also easy to represent the fundamental quantity F as the difference between the areas of two rectangles in the same figure. I

* Kasner, Proceedings of the National Academy, vol. 14 (1928), pp. 75–82.

have pointed out also that the diameter through O has its extremities at points γ_1 and γ_2 of the Kasner circle, so that $|\gamma_1|$ and $|\gamma_2|$ are precisely the maximum and the minimum values of $|\gamma|$, that is, the maximum and minimum values of the stretching ratio

$$R = r^{1/2} = |\gamma| = \left|\frac{dw}{dz}\right|,$$

discussed in §3. Finally, I have shown that the jacobian of $f(z)$ is represented by the length of the tangent from O to the Kasner circle. The jacobian of $f(z)$ can vanish, therefore, only for values of z for which the Kasner circle passes through the origin O in the γ plane. It follows that *the jacobian vanishes at a point z if and only if at least one determination of the directional derivative γ vanishes at that point.* This can be verified also immediately by substituting $\alpha=0$, $\beta=0$ in the equation (21).

9. *The Areal Derivative.* By Green's Theorem, we may write

$$(26)\qquad \int_C f(z)dz = \int_C (u+iv)(dx+idy) = 2i\iint_R \mathcal{P}[f(z)]dxdy,$$

where C is a closed curve, and where the double integral is taken over the region R bounded by C. If σ denotes the area of R, we have, by the law of the mean,

$$(27)\qquad \int_C f(z)dz = 2i\mathcal{P}[f(\zeta)]\iint_R dxdy = 2i\mathcal{P}[f(\zeta)]\cdot\sigma,$$

where ζ is some value of z in R. If C shrinks toward a single point z, ζ also approaches z, and we may write

$$(28)\qquad 2i\mathcal{P}[f(z)] = \lim \frac{1}{\sigma}\int_C f(z)dz.$$

This expression is defined by Pompeiu* to be the areal derivative.

If we introduce the conjugate imaginary quantity $\bar{z}=x-iy$, we may find, similarly,

* Pompeiu, Rendiconti di Palermo, vol. 33 (1912), p. 112. I have used the symbol $\mathcal{P}$ because it is his initial.

$$2i\mathfrak{D}[f(z)] = \lim \frac{1}{\sigma} \int_C f(z)d\bar{z}. \tag{29}$$

Hence the formula (5) may be written in the form

$$\gamma = \frac{dw}{dz} = -\frac{i}{2} \lim \frac{1}{\sigma}\left[\int_C f(z)d\bar{z} + e^{-2i\theta}\int_C f(z)dz\right]. \tag{30}$$

It may be remarked that the forms (28), (29), (30), and hence also the forms (26) and (27), can be interpreted even when it is *not* supposed by hypothesis that the first derivatives of u and v exist. Some results exist in this field, but they are not important. I propose, however, to pursue this and to investigate results obtainable without the hypothesis in question. It is to be remarked that the results of §10, and the well known theorem of Goursat mentioned there, may be construed to be theorems of this category.

Pompeiu, Hayashi, and Nicolesco* have considered at length the values of $\int f(z)dz$ on a *circle* of radius ρ about the point z_0; and have derived formulas for the expansion of this integral in terms of successive powers of ρ^2. From these expansions, several theorems are deducible. Thus, if $(1/(\pi\rho^2))\int f(z)dz$, taken over concentric circles of radius ρ, is *independent* of ρ, the function is called by Nicolesco *non-analytic of the first degree.* It is shown by Hayashi that for such a function $\nabla^2 u + i\nabla^2 v$ is itself an analytic function, where ∇^2 denotes the usual Laplace operator. In §10, we shall see that if the same expression approaches zero with ρ, for a set of values of z that is everywhere dense, then $f(z)$ is an analytic function.

The introduction of the conjugate imaginary $\bar{z}$ may be employed to simplify the notations as follows. Using the simple identities

$$x = \frac{z+\bar{z}}{2}, \qquad y = \frac{z-\bar{z}}{2i},$$

we may write

$$\begin{aligned} w = f(z) &= \phi\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i}\right) + i\psi\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i}\right) \\ &= F(z, \bar{z}), \end{aligned} \tag{31}$$

* Pompeiu, loc. cit.; Hayashi, loc. cit.; Nicolesco, Thesis, loc. cit.

whence, if we agree to differentiate *partially** with regard to z and $\bar{z}$, we may write, by an easy identification with (6),

$$\frac{\partial w}{\partial z} = \mathcal{D}(w), \quad \frac{\partial w}{\partial \bar{z}} = \mathcal{P}(w), \tag{32}$$

and†

$$\gamma = \frac{dw}{dz} = \frac{\partial w}{\partial z} + e^{-2i\theta}\frac{\partial w}{\partial \bar{z}}. \tag{33}$$

In order to refer by some name to $\mathcal{P}(w)$, I shall call it the *conjugate derivative*, for the reason suggested by (32), and because the name is otherwise appropriate; whenever it is convenient to do so, however, I shall use the combination $2i\mathcal{P}(w)$, which is the Pompeiu *areal derivative*. Similarly, I shall call the *integral* $\iint f(z)d\bar{z}$ the *conjugate integral*; for the most part, however, I shall use the *mean derivative* $\mathcal{D}(w)$, which, by (29), differs from this conjugate integral only by the factor $2i$.

Pompeiu considered the integral

$$\int_C \frac{f(z)}{z-\zeta}dz,$$

where ζ is any point in R; and he obtained the formula

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(s)}{s-z}ds + \frac{1}{\pi}\iint_R \frac{\mathcal{P}[f(s)]}{z-s}dxdy, \tag{34}$$

which is the generalization of the well known Cauchy Integral Theorem. Essentially the same formula is used by Borel in his *Leçons sur les Fonctions Monogènes*, and by Calugaréano in his thesis, where a complete proof is given. This formula shows that a function $f(z)$ is completely determined in R if its values are known on C, *and if the areal derivative is known in* R. The classical theorem is a special case of this one, since in the classical case the areal derivative is known to be identically zero.

The first term of (34), as is well known in the classical theory,

---

* Calugaréano, Thesis, p. 5, credits this suggestion to Al. Proca, who has himself not published it; such forms are used also by Nicolesco in his thesis.

† We may even write, with Calugaréano, $dw/dz = \partial w/\partial z + (\partial w/\partial \bar{z})\,(d\bar{z}/dz)$, since it is true that $d\bar{z}/dz = e^{-2i\theta}$.

represents a function analytic in R. It follows that *the knowledge of the areal derivative in R determines a function $f(z)$ to within an analytic function.* This follows also from the fact that if two functions f_1 and f_2 have the same areal derivative, the areal derivative of their difference is zero; that is, their difference is an analytic function. Calugaréano shows in his thesis (pp. 15–20) that the areal derivative can be assigned arbitrarily; that is, if $\mathcal{P}(w)$ be any function of z, $f(z)$ given by the second term of (34) has $\mathcal{P}(w)$ as its areal derivative.

Since the term $e^{-2i\theta}\mathcal{P}(w)$ in (5) essentially determines what Kasner has called a *clock*, the same theorems hold, by mere rewording, in terms of the concept of *clock*.

10. *The Morera Theorem.* The original Morera Theorem is contained in all standard works on the classical theory of functions. Refinements of the hypotheses have been made by several authors. H. Looman* has proved the following theorem.

If z is a point of a region R, and γ is a square containing z, consider the function

$$\sigma(z) = \limsup_{\gamma \to 0} \frac{\int_\gamma f(z)dz}{\gamma}, \tag{35}$$

where γ denotes also the area of the square; if then $\sigma(z)$ is everywhere finite and equal to zero almost everywhere, the function $f(z)$ is analytic throughout R.

A further extension of this theorem was given by J. Wolff.† These theorems may be regarded as extensions of the well known proof by Goursat‡ of the Cauchy Integral Theorem.

If we assume that the limit contained in Looman's theorem exists, we have in it again essentially the areal derivative of Pompeiu, so that $\sigma(z)$ is in that case equal to the areal derivative. With the assumptions regarding continuity made in the present paper, we may state the following theorem.

With the notation and assumptions of §2, if

---

* Nieuw Archief voor Wiskunde, vol. 14 (1924), pp. 234–239.

† Report of the Toronto Congress, pp. 457–459.

‡ Transactions of this Society, vol. 1 (1900), p. 14.

$$\mathcal{P}(w) = \lim \frac{i}{2\gamma} \int_{\gamma} f(z)dz$$

is zero at any set of points everywhere dense in R, then f(z) is analytic throughout R.

As a corollary, we may state the following generalization of the Morera Theorem.

With the notation and assumptions of §2, if $I_i = \int_{C_i} f(z)dz$ *approaches zero with i on a set of curves* C_i *that approach z, for each of a set of points that is everywhere dense in R, then f(z) is analytic throughout R.*

11. *The Liouville Theorem.* I have pointed out* that the familiar theorem that the absolute value of an analytic function cannot have a maximum in any region in which it is regular, can be extended to the general case of §2. For if we consider simply the real transformation (2), and if (u_0, v_0) corresponds to $x=0$, $y=0$, the existence of a maximum of $|f(z)| = u^2+v^2$ at any point (u_1, v_1) would imply that the equations (2) could not be solved to give a unique solution

$$x = \Phi(u, v), \quad y = \Psi(u, v),$$

in a region about (u_1, v_1). By the usual theorems on the existence of implicit functions, this would imply that the jacobian of (2) vanishes at (u_1, v_1). If we exclude the vanishing of the jacobian (which corresponds to a requirement that the derivative shall not vanish), the maximum cannot exist. We may therefore state the following theorem.

If the jacobian of (2) *does not vanish in a region R,* $|f(z)|$ *cannot be at a maximum at any point of R.*

Since, as I have pointed out in §8, the jacobian cannot vanish unless at least one value of the directional derivative at the same point is zero, and conversely, we may re-state this theorem as follows.

If the directional derivative γ *does not vanish in a region R,* $|f(z)|$ *cannot be at a maximum at any point of R.*

We may use this result to generalize the classical theorem of Liouville, that every integral analytic function whose absolute

* This Bulletin, vol. 36, p. 59, and vol. 36, p. 801.

value is always less than a fixed number M is a constant. For, suppose that any function $f(z)$ is defined, and that the assumptions of §2 hold, for every point z in the z plane, that $|f(z)| < M$ for all values of z, and suppose that $\lim_{z\to\infty} f(z) = 0$. Then $|f(z)|$ would have a maximum at the same point in the plane, which contradicts the preceding theorem, except in the trivial case $f(z) \equiv 0$. Next, suppose that $\lim_{z\to\infty} f(z) = C$; then it is sufficient to consider the new function $\phi(z) = f(z) - C$, for which $\lim_{z\to\infty} \phi(z) = 0$. Hence we may state the following theorem.

If $f(z)$ satisfies the assumptions of §2 at every point of the z plane, if the directional derivative is never zero, if $|f(z)| < M$, and if $\lim_{z\to\infty} f(z)$ exists, then $f(z)$ is a constant.

As in the classical theory, it is easy to define the point at infinity, by means of the simple transformation $z = 1/z'$. In the case just discussed, that is, when $\lim_{z\to\infty} f(z) = C$, the function $f(1/z')$ is regular near the new origin $z' = 0$ if we define it to be equal to C at that new origin. We may also use the classical language, and say that $f(z)$ has the value C at the point at infinity, and that it is defined at every point on the "sphere," in the usual sense. Since the theorems announced above are also valid near the origin in the z' plane, we may state the following theorem, which is a generalization of one form of the Liouville theorem.

If $f(z)$ satisfies the assumptions of §2 at every point of the z sphere, and if the directional derivative is never zero, then $f(z)$ is a constant.

It is possible to replace the condition that $\lim_{z\to\infty} f(z)$ exist by lighter requirements. If $|f(z)|$ assumes the value M at any point, it is easy to see that the theorem regarding the non-existence of a maximum would have to be violated at that point, if we maintain the assumption that the directional derivative never vanishes. It is indeed sufficient to assume that there exists some closed curve inside of which $f(z)$ takes on values greater than any values assumed on the curve itself.

Somewhat similar theorems may be found also by using the formula (34) in a manner analogous to the use made of the analogous theorem in the classical theory, for special hypotheses regarding the second term, that is, regarding the areal deriva-

tive. This is precisely what is done in the classical theory, for it is assumed there that the areal derivative is identically zero. If we assume that $|f(z)|$ has a maximum M at z_0, we may take C to be a circle of radius ρ about z_0, and we can then show from (34) that we must have

$$
\begin{aligned}
|f(z_0)| &\leqq \frac{1}{2\pi}\int_0^{2\pi} |f(s)|\, d\theta + \left|\frac{\mathcal{P}[f(\zeta)]}{\pi}\right| \iint d\rho d\theta \\
&\leqq \{\text{Max. of } |f(z)| \text{ on } C\} + 2\rho\mathcal{P}[f(\zeta)],
\end{aligned} \tag{36}
$$

where ζ is a value of z within C.

12. *The Derivative on a Field. Stieltjes Integrals.* Nicolesco in his thesis (pp. 21–25) has proposed the idea of the derivative of $f(z)$ on a given field. This field may be given by means of a family of curves

$$x = \phi(t, \alpha), \quad y = \psi(t, \alpha), \tag{37}$$

of which one and only one passes through a given point of R, or by a given real function $m(x, y)$ which determines a direction m at every point of R; that is, essentially by means of a differential equation. We shall assume the essential equivalence of these two methods of determining the field. Nicolesco has proved the following theorem.

If $m(x, y)$ is given, and if any function $\gamma(z) = \gamma = \alpha + i\beta$ is given, there exist an infinite number of functions $f(z) = u + iv$ whose derivatives on the field m are equal to $\gamma(z)$; and these are the solutions of the equations

$$
\begin{aligned}
\frac{\partial u}{\partial x} + m\frac{\partial u}{\partial y} - \alpha + m\beta = 0, \\
\frac{\partial v}{\partial x} + m\frac{\partial v}{\partial y} - m\alpha - \beta = 0,
\end{aligned} \tag{38}
$$

in which the variables u and v are separated.

The functions $f(z)$ may be called the primitive functions of the function $\gamma(z)$.

I have pointed out elsewhere* that integration and differentiation can be carried out on such a field, and that they are in-

* This Bulletin, vol. 34 (1928), p. 150; and Bulletin of the Calcutta Mathematical Society, vol. 20 (1928), pp. 109–124.

verse operations in the usual sense, if the directions taken in each case lie in the field. In fact, if (37) define the directions in question, $dx/dt=\cos\theta$ and $dy/dt=\sin\theta$ are determined at every point, so that $\theta=\theta(x, y)$ is a function of x and y. Then, given $w=f(z)=u(x, y)+iv(x, y)$,

$$(39)\quad \gamma=\left.\frac{dw}{dz}\right|_{\theta}=\alpha+i\beta=\frac{(u_x+iv_x)\cos\theta+(u_y+iv_y)\sin\theta}{\cos\theta+i\sin\theta}.$$

The integral, which I have called a *Stieltjes integral*, of this function γ on the curves of the family (37) is

$$\begin{aligned}\int_{z_0}^{z}\gamma dz&=\int_{x_0,y_0}^{x,y}(\alpha+i\beta)(dx+idy)\\&=\int(\alpha+i\beta)\{d\phi(t,\alpha)+id\psi(t,\alpha)\}\\(40)\qquad&=\int(\alpha+i\beta)(\cos\theta+i\sin\theta)d\theta\\&=\int\{(u_x+iv_x)\cos\theta+(u_y+iv_y)\sin\theta\}d\theta\\&=\int(u_x+iv_x)dx+(u_y+iv_y)dy=[u+iv]_{x_0,y_0}^{x,y}\end{aligned}$$

where the limits to be taken in the several integrals are obvious.

Conversely, the derivative on the field of the integral on the field is readily seen to be equal to the original function. For a given field, therefore, differentiation and integration may be regarded as inverse operations, the result in each case being a function of z alone.

13. *Second and Higher Derivatives.* The definitions of second and higher derivatives that have been given by different authors are not in agreement. In taking a second derivative, if the slope of the curve along which the increment Δz approaches zero be denoted by m', while m still denotes the similar slope for the first derivative, the following plans may be (and have been) used:

(A) $m=m'=$const., independent of z.

(B) $m=m'=\mu(x, y)$, a field (see §12).

(C) m', and the curve of approach, independent of m.

In his thesis, Nicolesco considers the various possibilities. First of all, he proves that in any case the order of differentiation is immaterial; that is, that m and m' may be interchanged without affecting the second derivative. He extends this result to the nth derivative. His position usually, however, is that of the assumption (A) above.

Calugaréano, who makes much more extensive use of second and higher derivatives, makes the assumption (A) throughout. He discusses the difference between his own attitude and that of Kasner in a recent paper.* We shall return to this discussion. Calugaréano proceeds with essentially the same notation as that used in equations (31), (32), (33) of §9 of the present paper. Assuming $m = m'$, we easily derive the formula

$$\frac{d^n w}{dz^n} = \left(\frac{\partial}{\partial z} + e^{-2i\theta}\frac{\partial}{\partial \bar{z}}\right)^{(n)} w, \tag{41}$$

where the symbolic exponent (n) is to be interpreted in the usual symbolic sense. The simplicity of this formula makes unnecessary any extended discussion, if this interpretation is to be adopted.

Kasner† and his students‡ have adopted the assumption (C). He uses as the path of approach any curve whatever, independent of the original choice of m, and he obtains the formula

$$\frac{d^2 w}{dz^2} = \frac{w_{xx} + 2w_{xy}y' + w_{yy}y'^2}{(1 + iy')^2} + \frac{w_y - iw_x}{(1 + iy')^3}\, y''. \tag{42}$$

This formula depends essentially upon the second derivative y'' of the curve of approach, that is, upon the curvature of that curve. He shows that d^2w/dz^2 is independent of y'', under his assumptions, when and only when w is an analytic function of z. He also proves that the ∞^1 set of values of d^2w/dz^2, for a constant curvature κ, describe an irrational curve of the eighth order, which reduces to a limaçon when $\kappa = 0$. A detailed report upon these facts seems inappropriate here, though the geometry developed is of great interest. Further developments of this

* Transactions of this Society, vol. 32 (1930), pp. 110–113.

† Transactions of this Society, vol. 30 (1928), pp. 803–818.

‡ Kasner and Hofmann, this Bulletin, vol. 34 (1928), p. 495; J. E. Donahue, Columbia University Thesis, Hamburg, 1930.

geometry have been given by J. E. Donahue in his thesis (loc. cit.). He includes a review of Kasner's work (only), and gives elaborate details with many special cases, and with very interesting geometric developments. These seem, however, to be beyond the scope of this paper.

14. *Solutions of Differential Equations.* Calugaréano, using the definitions stated here in §13, has proceeded to discuss the non-analytic solutions of differential equations and of functional equations. It will not be possible here, in the space at my command, to do more than to indicate the nature of his results, and to state a few of the simplest theorems. It may be said that these developments would not be possible under the definitions assumed by Kasner, and that it would seem that differential equations of the sort discussed below would not possess non-analytic solutions if the Kasner definition were to be used.

Returning to the definition used by Calugaréano, we may say that a differential equation

$$\Phi(z, w, w', \cdots, w^{(n)}) = 0 \tag{43}$$

will be satisfied, in general, by a function $w = f(z)$ if, when we substitute for the nth derivative $w^{(n)}$ the value given in (41), the equation (43) is satisfied for all values of θ. An attempt to express the fact that the result of the substitution is independent of θ gives in general a number of conditions which must be satisfied if there are to be any non-analytic solutions. That non-analytic solutions do actually exist in some cases is easy to verify. Thus the linear equation

$$A(z)w'' + B(z)w' + C(z)w + D(z) = 0, \tag{44}$$

where A, B, C, D are analytic, has solutions if

$$B^2 - 4AC = 2(A'B - AB'), \tag{45}$$

a condition which is fulfilled for any A and B if C is properly chosen, provided A does not vanish.

Such linear equations, however, always have rather simple non-analytic solutions; if we express them in the form $F(z, \bar{z})$, as in (31), it results that F is an analytic function of z *and a polynomial in* $\bar{z}$, for linear equations of any order. However, these solutions are helpful also in obtaining ordinary analytic solu-

tions, since it is shown that the analytic function $F(z, z-C)$ obtained by replacing $\bar{z}$ by $z-C$, is also a solution of the given differential equation.

Calugaréano also showed that any differential equation of the second order that has non-analytic solutions must be of the form

$$w'' + A(z, w)w'^2 + B(z, w)w' + C(z, w) = 0, \tag{46}$$

where A, B, and C are analytic functions of z and w. It is true in this case also that if $F(z, \bar{z})$ is a solution of (46), then $F(z, z-C)$ is an analytic solution of the same equation. These results are generalized to equations of higher order.

In his 1930 paper cited above, Calugaréano shows that the analytic integral of (46) is obtained from the non-analytic integral $F(z, \bar{z})$ by replacing $\bar{z}$ by $\gamma_1 z+\gamma_2$.

15. *Other Results. Conclusion.* The results mentioned above have explicit reference to theorems on non-analytic functions. Two other directions exist in which associated results are to be found. In the first place, as is emphasized above, even the classical textbooks contain many theorems in which the hypothesis of analytic character is not made. A complete report on theorems satisfied by non-analytic functions would have to include all such theorems, which would not be appropriate here. Moreover, as the theory of non-analytic functions is becoming more widely known, more attention has been and will be paid to the question as to whether or not the hypothesis of analytic character is necessary, in any given theorem. I have called attention to several theorems in which such a hypothesis can be modified without essential change in the conclusion of the theorem. I have not tried to include all such instances in this paper. It is to be expected that a very large number of such cases will be found, as has been true in the similar case in the theory of functions of real variables.

Other notable omissions are caused through lack of space. In a list which follows, I have included the titles of some papers to which I have made no explicit references. In general, I have sought to include results which seemed to possess not only interest, but also a degree of apparent finality. Some results appear to be tentative in character, as is natural in a new field.

The other direction of research associated with this paper, but not represented here, is the generalization of ideas inherent in the background of this paper, through extension to higher dimensional spaces, or to functions of other fundamental elements than points, or to hypercomplex numbers. A number of papers of this character have been published, and they are highly significant, but they seemed not to be an integral part of the discussion here, unless the field of discussion were to be extended unduly, in view of the limitation of space. That I myself have assisted in the publication of some of these is evidence enough that my present omission of them is not caused by any lack of personal interest.

It is my hope that the present paper may serve to acquaint many with some of the salient advances in the theory, to enable others to find a variety of results in one compact paper, and to give some indication of possible future research.

BIBLIOGRAPHY

G. Calugaréano, *Sur les fonctions polygènes d'une variable complexe*, Thesis, Paris, 1928.

G. Calugaréano, Comptes Rendus, vol. 186 (1928), p. 186; vol. 186 (1928), p. 930; vol. 186 (1928), p. 1406.

G. Calugaréano, Bulletin de la Société des Sciences de Cluj, vol. 14 (1928), pp. 357–361.

G. Calugaréano, Transactions of this Society, vol. 31 (1929), p. 372; vol. 32 (1930), p. 110.

J. E. Donahue, *Concerning the geometry of the second derivative of a polygenic function*, Columbia University Thesis, Hamburg, 1930.

E. R. Hedrick, L. Ingold, and W. D. A. Westfall, *Theory of non-analytic functions of a complex variable*, Journal de Mathématiques, (9), vol. 2 (1923), pp. 327–342.

E. R. Hedrick and L. Ingold, *Analytic functions in three dimensions*, Transactions of this Society, vol. 27 (1925), pp. 551–555.

E. R. Hedrick and L. Ingold, *The Beltrami equations in three dimensions*, Transactions of this Society, vol. 27 (1925), pp. 556–562.

E. R. Hedrick, *On the derivatives of non-analytic functions*, Proceedings of the National Academy, vol. 14 (1928), pp. 649–654.

E. R. Hedrick, *On certain properties of non-analytic functions of a complex variable*, Bulletin of the Calcutta Mathematical Society, vol. 20 (1928), pp. 109–124; see also this Bulletin, vol. 34 (1928), pp. 150, 435–6, 703; vol. 35 (1929), pp. 176, 599; vol. 36 (1930), pp. 59, 368, 801.

Lulu Hofmann and E. Kasner, *Homographic circles or clocks*, this Bulletin, vol. 34 (1928), pp. 495–503.

E. Kasner, *A new theory of polygenic (or non-monogenic) functions*, Science, vol. 46 (1927), pp. 581–582.

E. Kasner, *General theory of polygenic (or non-monogenic) functions. The derivative congruence*, Proceedings of the National Academy, vol. 13 (1928), pp. 75–82.

E. Kasner, *Géométrie des fonctions polygènes*, Atti del Congresso, Bologna, 1928, vol. 6, pp. 255–260.

E. Kasner, *The second derivative of a polygenic function*, Transactions of this Society, vol. 30 (1928), pp. 803–816.

E. Kasner, *Note on the derivative circular congruence of a polygenic function*, this Bulletin, vol. 24 (1928), pp. 561–565.

P. W. Ketchum, *Analytic functions of hypercomplex variables*, Transactions of this Society, vol. 30 (1928), pp. 641–667.

P. W. Ketchum and Ted Martin, *Polygenic functions of hypercomplex variables*, this Bulletin, vol. 38 (1932), pp. 66–72.

Edna E. Kramer, *Polygenic functions of the dual variable* $w=u+jv$, Columbia University Thesis, 1930; see also American Journal of Mathematics, vol. 52 (1930), pp. 370–376.

L. Lichtenstein, Sitzungsberichte, Berliner Mathematische Gesellschaft, 1910, pp. 84–100.

H. Looman, *Ueber eine Erweiterung des Cauchy-Goursatschen Integralsatzes*, Nieuw Achief voor Wiskunde, vol. 14 (1924), pp. 234–239.

P. Montel, Comptes Rendus, vol. 156 (1913), pp. 1820–1822; Annales de l'École Normale, (3), vol. 24 (1907), pp. 283–295.

M. Nicolesco, *Fonctions complexes dans le plan et dans l'espace*, Thesis, Paris, 1928; Comptes Rendus, vol. 185 (1927), p. 442.

D. Pompeiu, *Sur la continuité des fonctions d'une variable complexe*, Thesis, Paris, 1905.

D. Pompeiu, *Sur une classe de fonctions d'une variable complexe*, Rendiconti di Palermo, vol. 23 (1912), pp. 108–113; vol. 35 (1913), p. 277.

V. C. Poor, *Residues of polygenic functions*, Transactions of this Society, vol. 32 (1930), pp. 216–222.

H. Rademacher, *Bemerkungen zu den Cauchy-Riemannschen Differentialgleichungen und zum Moreraschen Satz*, Mathematische Zeitschrift, vol. 3 (1919).

G. E. Raynor, *Generalization of the Beltrami equations to curved n-space*, this Bulletin, vol. 33 (1927), pp. 435–439.

J. Ridder, *Ueber den Cauchyschen Integralsatz für reelle und komplexe Funktionen*, Mathematische Annalen, vol. 102 (1929), pp. 132–156.

J. Ridder, *Ein Satz über iterierte Integrale und seine Anwendung zur Untersuchung der Analyticität komplexer Funktionen*, Mathematische Zeitschrift, vol. 31 (1929), pp. 141–148.

J. Ridder, *Ueber Analyticität komplexer Funktionen*, Nieuw Archiv voor Wiskunde, (2), vol. 16 (1929), pp. 17–19.

J. Wolff, *On the sufficient conditions for analyticity of functions of a complex variable*, Report of the Toronto Congress, pp. 457–459; see also Nieuw Archiv voor Wiskunde, vol. 14 (1925),

THE UNIVERSITY OF CALIFORNIA AT LOS ANGELES

SETS OF LOCAL SEPARATING POINTS OF A CONTINUUM*

BY G. T. WHYBURN

1. *Introduction.* Let M denote any locally compact metric continuum and let L be the set of all local separating points† of M. We proceed to establish the following six properties, of which, for our immediate purposes, the most useful is number (iv).

(i). *If U is any uncountable subset of L, there exists a point x of U which is a point of order 2 in M relative to U.*

This statement means that x is contained in arbitrarily small neighborhoods whose boundaries have in common with M just two points and these two points belong to U. A proof has already been given by the author (loc. cit.).

(ii). *If H is any connected subset of M, then $(\overline{H}-H)\cdot L$ is countable.*

For if not, (i) would give a point x of this set which could be separated in M from some point of H by two points not in H, which obviously is impossible since $H+x$ is connected.

(iii). *If H is any connected subset of M, the points of $H\cdot L$ which are not local separating points of H are countable.*

This results immediately from (i).

(iv). *If H is any connected subset of M such that $\overline{H} \subset L+C$, where C is some countable set, then H is a locally connected G_δ-set. Hence H is arcwise connected.*

By (ii) we see that $(\overline{H}-H)\cdot L$ and hence $\overline{H}-H$ itself is countable. Thus H is a G_δ-set. Now $\overline{H}$ must be a regular curve, for by (i), all save a countable number of its points are points of order 2. Thus any connected subset of $\overline{H}$, and in particular H, is locally connected. That H is arcwise connected follows now by the well known theorem of Moore-Menger.‡

---

* Presented to the Society, February 25, 1933.

† A point p is a local separating point of M provided some neighborhood V of p exists such that $M\cdot\overline{V}-p$ is separated between some pair of points belonging to the component of $M\cdot\overline{V}$ which contains p. See the author's paper in Monatshefte für Mathematik und Physik, vol. 36 (1929), pp. 305–314.

‡ See R. L. Moore, *Foundations of Point Set Theory*, Colloquium Publications of this Society, vol. 13 (1932), p. 86; and K. Menger, Monatshefte für Mathematik und Physik, vol. 36 (1929), pp. 193–218.

(v). *If N is any continuum $\subset L+C$, where C is countable, then every connected subset of N is a locally connected G_δ-set and hence is arcwise connected.*

This is a corollary to (iv).

(vi). *If M is locally connected, then every connected subset H of L is the difference between an F_σ and a countable set.*

For it is known* that in this case L is an F_σ, so that $M-L$ is a G_δ and hence so also is $(M-L)\cdot\overline{H}=(M-L)\cdot(\overline{H}-H)$. Whence, $H=[\overline{H}-(M-L)\cdot(\overline{H}-H)]-L\cdot(\overline{H}-H)$, and the first of these two sets is an F_σ and the second, by (ii), is countable.

2. THEOREM. *In order that every connected subset of a continuum M be a G_δ it is necessary and sufficient that the set N of non-local-separating points of M be countable.*

The sufficiency of the condition results immediately from (iv), in view of the fact that, for any connected subset H of M, we have $\overline{H}\subset M=L+N$, and N is countable.

To prove that the condition is necessary, we suppose, on the contrary, that N is uncountable. Let $M=M_1+M_2$, where M_1 and M_2 are disjoint and totally imperfect† and where M_2, say, contains uncountably many points of N. Then‡ M_1+L is connected and $M-(M_1+L)=E$ is totally imperfect and uncountable. Thus E is not an F_σ and hence M_1+L is not a G_δ.

COROLLARY 1. *If every connected subset of M is a G_δ, then M is a regular curve, no cyclic element of M has a continuum of condensation, and the end points of M are countable.*

COROLLARY 2. *If all save a countable number of the points of each cyclic element C of a locally connected continuum M are local separating points of C, then every connected subset of M is arcwise connected.*

COROLLARY 3. *If the non-local-separating points of each cyclic element of a locally connected continuum M are countable, then every connected subset of M will be a G_δ if and only if the end points of M are countable.*

---

* See the author's paper, Mathematische Annalen, vol. 162 (1929), p. 318.

† That is, neither contains any perfect set. For a proof that such a division of M is possible, see F. Bernstein, Leipziger Berichte, vol. 60 (1908), p. 325 and Hausdorff, *Mengenlehre*, 1927, p. 156.

‡ See my paper *On the existence of totally imperfect sets . . .*, American Journal of Mathematics, vol. 55 (1933), pp. 146–152.

For any local separating point of a cyclic element of M is a local separating point of M and any non-local-separating point of M which is on no non-degenerate cyclic element of M must be an end point of M.

3. THEOREM. *If M is any locally compact continuum such that* (a) *no two maximal free arcs* in M abut and* (b) $L \cdot \overline{R}$ *is countable and* $R \neq 0$, *where R is the set of all ramification points (that is, points of order >2) of M, then M contains a connected subset which is not arcwise connected.*

Proof. Let $M = M_1 + M_2$, where M_1 and M_2 are totally imperfect and disjoint. Set $E = M_1 + L$. Then since $M - E \subset M_2$, E is connected (loc. cit.). Furthermore $E \cdot \overline{R} = M_1 \cdot \overline{R} + L \cdot \overline{R}$, and since $L \cdot \overline{R}$ is countable it follows that $E \cdot \overline{R}$ is totally imperfect. Now if a and b are two points of E lying in different maximal free arcs† of M, there can exist no arc ab in E. For if ab is any arc in M from a to n, then $ab \cdot \overline{R}$ cannot be countable, (for if so, some two free arcs contained in ab would abut), and hence it must contain a perfect set. Thus ab cannot be $\subset E$, since $E \cdot \overline{R}$ is totally imperfect.

4. EXAMPLE. *There exists a regular curve C such that* (a) *no two free arcs of C abut,* (b) $L \cdot \overline{R}$ *is countable, and* (c) $\overline{R}$ *is punctiform. Hence C contains a connected subset which is not arcwise connected.*

Let I be the unit interval and let K be the non-dense perfect set consisting of all numbers on I which can be expressed in the triadic number system using only the digits 0 and 2. Let I_1, I_2, $\cdots$, be the segments on I complementary to K ordered in descending order of length and let $p_1, p_2, \cdots$, be the end points of these segments, where p_{2n-1} and p_{2n} are the end points of I_n. Let $a_n = \sum_1^n 2^m$ and $P_n = \sum_1^{a_n} p_i$ and $P = \sum_1^\infty P_i$. Let us select in $K - P$ two sequences of points x_n and y_n such that

$$0 \leftarrow \cdots < x_n < x_{n-1} < \cdots < x_1 < y_1 < \cdots < y_{n-1} < y_n < \cdots \rightarrow 1.$$

* An arc ab is *free* in M provided $ab - (a+b)$ is an open subset of M. Two such arcs are said to abut if they have a common end point.

† Clearly two such exist, otherwise M contains a punctiform connected set. See my paper, loc. cit.

In the upper half plane let us construct equilateral triangles with bases x_2y_2, $y_{4n-1}\,y_{4n+2}$, $x_{4n+2}x_{4n-1}$, $(n=1, 2, 3, \cdots)$, and in the lower half plane construct equilateral triangles with bases $y_{4n-3}y_{4n}$ and $x_{4n}x_{4n-3}$, $(n=1, 2, 3, \cdots)$. Let T_0 be the set of triangles so constructed.

Now the points (x_n) and (y_n) together with the points of P_1 divide I into a collection of intervals; let us omit I_1 from this collection and call J_1 the resulting collection. Now on each interval of J_1 let us construct a set of equilateral triangles exactly as we constructed T_0 on I. Let T_1 be the set of triangles obtained for all intervals of J_1. Now the vertices of the triangles of T_1 that are on I together with the points of P_2 divide I into a set of intervals. Omit from this set the intervals I_1, I_2, I_3 and call J_2 the resulting collection. On each interval of J_2 construct a set of triangles as before and call T_2 the total set so constructed.

Continue this process indefinitely and let $C=I+\sum_0^\infty T_n$. Then clearly C is a continuum. The maximal free arcs in C are exactly the intervals I_n together with the two sides of each of the triangles of $[T_n]$ not on I, so that no two maximal free arcs in C abut. Furthermore the set R of ramification points of C is a subset of K such that $\overline{R}=K$. Thus $\overline{R}$ is totally disconnected and C is a regular curve. Finally we note that if x is any point of $K-(P+R)$, then for each n there is a triangle t of T_n with base axb, where $0<a<x<b<1$; and since $\delta(t)<1/(n+1)$ and x is not a ramification point, it follows that x is not a local separating point. Thus $L\cdot\overline{R}=L\cdot K\subset P+L\cdot R$, which shows that $L\cdot\overline{R}$ is countable, since both P and $L\cdot R$ are countable.

THE JOHNS HOPKINS UNIVERSITY

ON THE EXISTENCE OF THE ABSOLUTE MINIMUM IN PROBLEMS OF LAGRANGE†

BY L. M. GRAVES

In a previous paper,‡ I remarked that the theorems obtained are applicable to problems of Lagrange in which the side conditions are of the form

$$g_\mu(x, y_1, \cdots, y_k) = 0, \quad (\mu = 1, \cdots, m). \tag{1}$$

It is easily seen also that *the proofs* there made are applicable under suitably *weakened hypotheses* (to be stated below) to problems of Lagrange in which appear side conditions which are *differential equations* of the form

$$y_\sigma' = h_\sigma(x, y_1, \cdots, y_k), \quad (\sigma = 1, \cdots, s). \tag{2}$$

Calculus of variations problems in which the integrand involves higher derivatives are reducible to Lagrange problems of this type, as are also isoperimetric problems in which the integrands of the integrals to be kept constant depend only on the coordinates. Examples in which the stronger hypotheses of the paper cited do not hold are given in the final paragraph.

We suppose that the integrand function $f(x, y_1, \cdots, y_k, y_1', \cdots, y_k')$ and the functions g_μ, h_σ, together with the partial derivatives $f_{y_i'}$, $g_{\mu x}$, $g_{\mu y_i}$, are defined and continuous for all points (x, y) in a closed domain A and for all y'. Let R^* denote the set of all points (x, y, y') having (x, y) in A and satisfying equations (1), (2) and

$$g_{\mu x} + g_{\mu y_i} y_i' = 0, \qquad (\mu = 1, \cdots, m). \tag{3}$$

An *admissible curve* C, $y_i = y_i(x)$, is one which is absolutely continuous and has all its elements (x, y, y') in R^*.§ Then if K is a closed class of absolutely continuous curves, the sub-class K^*

---

† Presented to the Society, December 27, 1932.

‡ *On the existence of the absolute minimum in space problems of the calculus of variations*, Annals of Mathematics, vol. 28 (1927), pp. 153–170.

§ The set of points x at which one or more of the functions y_i fails to have a finite derivative may be neglected.

consisting of all admissible curves in K is either closed or empty.

We replace hypotheses (I), (II) and (III), pages 158 and 167 of the paper cited, by the following:

(I*) $E(x, y, y_0', y') \geqq 0$ for all points (x, y_0, y_0') and (x, y, y') in R^*.

(II*) For every bounded set A' contained in A there exist constants $M>1$ and $\alpha>1$ such that $f(x, y, y') > \|y'\|^\alpha$ for all sets (x, y, y') in R^* having (x, y) in A' and $\|y'\| > M$.

(III*) For every bounded set A' in A and every number $L>0$ there exists a number $\rho>0$ such that, if C is an admissible ordinary curve having at least one point in A' and at least one point outside the sphere S_ρ of radius ρ, then the integral

$$I_C = \int_C f(x, y, y')dx > L.$$

With these weakened hypotheses, the conclusions of Theorems 1, 2, 3 and 4 are still valid for classes of admissible curves.

Criteria which are sufficient to ensure the fulfillment of hypothesis (III*) may be selected from the following. In order to state these additional conditions it is convenient to change the notation. Let $t+s=k$, and

$$I = \int f(x, y_1, \cdots, y_t, z_1, \cdots, z_s, y_1', \cdots, y_t', z_1', \cdots, z_s')dx,$$

$$(1') \qquad g_\mu(x, y, z) = 0, \qquad (\mu = 1, \cdots, m < t),$$

$$(2') \qquad z_\sigma' = h_\sigma(x, y, z), \qquad (\sigma = 1, \cdots, s).$$

Let $\|yz\|$ = the greater of $\|y\|$, $\|z\|$.

(IV) The maximum difference of abscissas of points in the domain A is finite and equal to D.

(V*) The integrand $f(x, y, z, y', z')$ has a finite lower bound in the domain R^*.

(VI*) There exist constants $M>0$ and $\alpha \geqq 1$ and a function $\phi(y, z)$ continuous for all (y, z) such that

$$(i) \qquad \phi(y, z) > 0$$

for $\|y\| \geqq M$;

(ii) $$\lim_{\|yz\|\to\infty} \|yz\|^{\alpha}\phi(y, z) = +\infty;$$

(iii) $$f(x, y, z, y', z') \geqq \|y'\|^{\alpha}\phi(y, z)$$

for all (x, y, z, y', z') in R^* with $\|y\| \geqq M$, $\|y'\| \geqq M$.

(VII*) There exist positive constants c, d, α, β, such that

$$f(x, y, z, y', z') = \Phi(x, y') - \Psi(x, y, z, z'),$$

where

$$\Phi(x, y') \geqq c\|y'\|^{1+\alpha+\beta} - d, \quad \Psi(x, y, z, z') \leqq d\|yz\|^{1+\alpha} + d,$$

for all (x, y, z, y', z') in R^*.

(IX*) There exists a constant Γ such that on every admissible curve, and for every x and x_1,

$$\|z(x)\| \leqq \Gamma[\max \|y(\xi)\| + \|z(x_1)\| + 1].$$

(VI**) Same as (VI*), except that the function ϕ is independent of the z_σ, and properties (i) and (ii) do not involve the z_σ.

(IX**) For every number q there is a number Λ_q such that $|h_\sigma(x, y, z)| \leqq \Lambda_q$ for every (x, y, z, y', z') in R^* with $\|y\| \leqq q$.

Condition (III*) is implied by (a) (IV), (V*), (VI*), (IX*); or by (b) (IV), (V*), (VI**), (IX**); or by (c) (IV), (VII*), (IX*). For the proofs of these statements, I shall indicate the alterations necessary in the proofs given on pages 168–170 of the paper cited.

Let K^* be a class of admissible ordinary curves each of which has at least one point in the bounded set A'. Let $p > 1$ be greater than the upper bound of $\|yz\|$ in the set A'. Then if $\|y\| \leqq q > p$ on a curve of K^*, also $\|z\| \leqq 3\Gamma q$ if (IX*) holds, and $\|z\| \leqq \Lambda_q D + p$ if (IV) and (IX**) hold. In either case, if $\eta_C \equiv$ maximum $\|y\|$, $\zeta_C \equiv$ maximum $\|yz\|$, on a curve C of the class K^*, then η_C becomes infinite with ζ_C. Let C be an arc of the class K^* having maximum $\|y\| \equiv Y > 2p$, and let $\overline{C}$ be a sub-arc having $\|y\| = Y$ at one end point, $\|y\| = p$ at the other end point, and $p < \|y\| < Y$ at interior points. If the conditions of group (a) hold, let ϕ_q be the minimum of $\phi(y, z)$ for $p \leqq \|y\| \leqq q$, $\|z\| \leqq 3\Gamma q$. If the conditions of group (b) hold, let ϕ_q be the minimum of $\phi(y)$ for $p \leqq \|y\| \leqq q$.

A few examples follow.

(1) $f = (zy' - yz')^2 - x^2, \qquad g = x^2 + y^2 + z^2 - 1 = 0,$

the region A is that included between $x = e - 1$ and $x = 1 - e$, $0 < e < 1$.

(2) $$f = [d^k z/dx^k]^{2n},$$

where n and k are positive integers.

(3) $$f = \frac{y'^4}{1 + z^2}, \quad z' = y, \quad \text{or} \quad z' = (1 + y^2)^{1/2}.$$

(4) $$f = \frac{y'^2}{(1 + z^2)^{1/2}}, \quad z' = y, \quad \text{or} \quad z' = (1 + y^2)^{1/2}.$$

(5) $$f = \frac{y'^4(1 + z^2)}{1 + y^2}, \quad z' = e^y.$$

(6) $$f = y'^2 - y - z, \quad z' = (1 + y^2)^{1/2}.$$

(7) $$f = y'^2 - z'^2, \quad z' = h(x, y, z),$$

where h is a bounded function.

(8) $$f = y'^4 - z'^3, \quad z' = h(x, y, z),$$

where

$$| h(x, y, z) | < \Gamma[| y | + 1].$$

In each example, a minimum exists in the class of all admissible ordinary curves joining two fixed points, provided that class is not empty. In Example (2), we may suppose the values of z and its first $k - 1$ derivatives given at the end points. In examples (7) and (8), the function $h(x, y, z)$ should be substituted for z' in the integrand f before computing the Weierstrassian E-function.

THE UNIVERSITY OF CHICAGO

SOME NEW FACTORIZATIONS OF $2^n \pm 1$

BY D. H. LEHMER

1. *Introduction.* The purpose of this note is to announce five new factorizations of $2^n \pm 1$. The results are due directly and indirectly to the writer's new number theoretic machine.* In fact the large primes quoted below would have remained undiscovered by the writer had he not been in possession of the means of factoring numbers of this size without difficulty. The actual factorizations quoted below are the result of a month's experiment with the machine. As a matter of fact the machine ran only a few hours altogether. When some permanent location has been found for the apparatus, it is hoped to consider much more difficult problems.

2. *Theorem for Large Primes.* We begin with the large primes. These primes were identified with the help of the following theorem.†

THEOREM A. *If the integer N divides $\alpha^{N-1}-1$, but is prime to $\alpha^{(N-1)/p}-1$, p a prime, then all the factors of N are of the form $px+1$.*

If a large enough prime factor p of $N-1$ is known, this theorem so restricts the factors of N, that the primality of N follows almost at once.

3. *Factorization of $2^{73}+1$.* The number N in this case is

$$N = \frac{2^{73}+1}{3 \cdot 1753} = 1795\ 91803\ 87410\ 70627.$$

It was found that $N-1$ is divisible by $p=811$, and that the hypothesis of Theorem A is satisfied for $\alpha=3$. Hence every factor of N is of the form $811x+1$, as well as $73x+1$ and $8x+1$, 3. Writing $N=a^2-b^2$, we deduce at once

$$a = 28039961672k + 244363342366.$$

* This Bulletin, vol. 38 (1932), p. 635.

† This theorem is a special case of Theorem 3, of the writer's article published in this Bulletin, vol. 33 (1932), p. 331.

Since the smallest factor of N is known to exceed 300000, we find

$$a \leqq 2993196881235,$$

so that $k \leqq 107$. These few values of k are easily disposed of using small moduli $\leqq 19$. Hence N is a prime and we have

$$2^{73} + 1 = 3 \cdot 1753 \cdot 1795\ 91803\ 87410\ 70627.$$

4. *Factorization of* $2^{85}-1$. The number N under consideration here is

$$N = \frac{2^{85} - 1}{(2^{17} - 1)(2^5 - 1)} = 9520\ 97280\ 63337\ 58431.$$

Here $N-1$ is divisible* by $p=257$. The hypothesis of Theorem A is shown to be satisfied with $\alpha = 3$, and the primality of N follows easily as before. Hence

$$2^{85} - 1 = 31 \cdot 131071 \cdot 9520\ 97280\ 63337\ 58431.$$

5. *Factorization of* $2^{95}+1$. The number N in this case is

$$N = \frac{(2^{95} + 1)(2 + 1)}{(2^{19} + 1)(2^5 + 1)(2281)} = 3011\ 34747\ 96142\ 49131.$$

It was found that N divides $3^{N-1}-1$, but no large prime factor of $N-1$ was immediately available. In fact

$$N - 1 = 2 \cdot 3 \cdot 5 \cdot 19 \cdot 5\ 28306\ 57537\ 09209.$$

It thus became necessary to examine the large factor M of $N-1$. By expanding $M^{1/2}$ in a regular continued fraction it was found that -7 is a quadratic residue of M. Hence if M is composite we may expect at least two representations of M by the form x^2+7y^2. The machine gave them in less than half an hour as follows:

$$N = 40923451^2 + 7 \cdot 22704112^2 = 66855539^2 + 7 \cdot 10779628^2.$$

This gives us the decomposition

$$M = 59957 \cdot 88114244437.$$

This last factor we proved to be a prime by applying Theorem

* One may easily show in general that any prime factor of $a^{p-1}-1$ not dividing a^q-1, divides $[(a^{pq}-1)(a-1)]/[(a^p-1)(a^q-1)]-1$.

A using the factor $p=2489947$ of 88114244436. Armed with this large factor of $N-1$, the application of Theorem A to N was soon completed, and it resulted that N is a prime. Hence

$$2^{95}+1 = 3\cdot 11\cdot 2281\cdot 174763\cdot 3011\ 34747\ 96142\ 49131.$$

This proof for primality is a positive proof, although it may seem round about. The machine could have been easily set to search directly for the factors of N, but its failure to find a factor might not have been acceptable to the reader as a proof of primality.

6. *Factorization of* $2^{93}+1$. The number under discussion here is

$$N=\frac{(2^{93}+1)(2+1)}{(2^{31}+1)(2^3+1)} = 1537\ 22867\ 20933\ 01419.$$

Since each factor of N is of the form $93x+1$ and $8x+1$, 3 we find easily that if $N=a^2-b^2$, then a is of the form $69192k+1239854518$. The smallest factor of N exceeds 300000 so that

$$a \leqq 2562047936822, \text{ or } k \leqq 37010176.$$

The machine was set to exclude values of k. It ran only a few seconds, and delivered the value $k=6886$. This gives us

$$a = 1716310630, \text{ and } b = 1186799691.$$

Hence

$$N = 529510939\cdot 2903110321.$$

It is easily seen that both factors are primes, hence

$$2^{93}+1 = 3^2\cdot 529510939\cdot 715827883\cdot 2903110321.$$

7. *Factorization of* $2^{79}-1$. We are concerned in this case with

$$N = (2^{79}-1)/2687 = 2\ 24958\ 286426\ 02584\ 99201.$$

Since the factors of N are of the forms $79x+1$, $8x\pm 1$, it follows that if $N=a^2-b^2$, a will have the forms

$$a = 6241x+159,\ 3x+0,\ 512x\pm 63,\ 97,\ 159,\ 191.$$

Combining these forms we get

$$a = 9586176k+159,\ 1610337,\ 4793247,\ 6403425,$$
$$6590655,\ 7002561,\ 8987199,\ 9399105.$$

Since $a \leqq 374930473917097$, we have in each case $k \leqq 39111579$. Thus the problem of representing N as the difference of squares was split into 8 parts. The first two parts were covered by the machine without any result. On the third run, however, the machine stopped almost at once at $x = 58088$. This gives

$$a = 556846584735, \quad b = 556644555032.$$

Hence we have the factorization

$$2^{79} - 1 = 2687 \cdot 202029703 \cdot 1113491139767.$$

It is not difficult to show that the factors are primes. This is the 13th composite Mersenne number to be completely factored. The author's recent report* on Mersenne numbers should be changed accordingly.

PASADENA, CALIFORNIA

MATRICES WHOSE sTH COMPOUNDS ARE EQUAL

BY JOHN WILLIAMSON

If A is a matrix of m rows and n columns and s is any positive integer less than or equal to the smaller of n and m, from A can be formed a new matrix A_s of ${}_mC_s$ rows and ${}_nC_s$ columns, the elements in the tth row of A_s being the ${}_nC_s$ determinants of order s that can be formed from the t_1th, $\cdots$, t_sth rows of A, and the elements in the tth column being the ${}_mC_s$ determinants of order s that can be formed from the t_1th, $\cdots$, t_sth columns of A. The matrix A_s, so defined, is called the sth compound matrix of A. In the following note we discuss the necessary and sufficient conditions under which the sth compounds of two matrices are equal. We shall require the following lemmas.

LEMMA I. *The rank of the sth compound of a matrix A, whose rank is r, is ${}_rC_s$ if $r \geqq s$ and is zero if $s > r$.*†

---

* This Bulletin, vol. 38 (1932), p. 384. Dr. N. G. W. H. Beeger has kindly called my attention to the fact that $2^{223} - 1$ has two known prime factors and should be classified accordingly.

† Cullis, *Matrices and Determinoids*, vol. 1, p. 289.

LEMMA II. *The sth compound of the product of two matrices is the product of the sth compounds of the two matrices, or, in symbols,**

$$(AB)_s = A_s B_s. \tag{1}$$

THEOREM. *If A is a matrix of rank r, the necessary and sufficient condition that $A_s = B_s$ is that*

(a) *the rank of B be less than s when $r < s$;*

(b) *there exist two non-singular matrices C and D such that*

$$CAD = \begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix}, \quad CBD = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix},$$

where T and S are two non-singular matrices of r rows and columns such that $|T| = |S|$, when $r = s$;

(c) *$A = \omega B$, where ω is an sth root of unity, when $r > s$.*

In case (a) if $A_s = B_s$, then $B_s = 0$ and by Lemma I the rank of B is less than r. On the other hand if the rank of B is less than s, then $B_s = 0 = A_s$. In case (b) the sufficiency of the condition follows from (1) and the fact that

$$\begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix}_s = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}_s.$$

We now proceed to prove that the condition stated above is necessary. Since A has rank r there exist two non-singular matrices C and D such that

$$CAD = R = \begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix},$$

where T is any non-singular r-rowed square matrix. If

$$CBD = F = \begin{pmatrix} S & G \\ H & K \end{pmatrix},$$

where S is an r-rowed square matrix, G an r by $n-r$ matrix, H an $m-r$ by r matrix, and K an $m-r$ by $n-r$ matrix, then, since $A_s = B_s$, it follows that $R_s = F_s$ and $|S| = |T| \neq 0$. Since R_s contains only one element different from zero, every determinant of order s that can be formed from $s-1$ columns of S and one of G is zero. If

$$S = (s_{ij}), \quad G = (g_{iq}), \quad (i, j = 1, 2, \cdots r; \quad q = 1, 2, \cdots, n-r),$$

* H. W. Turnbull, *Determinants, Matrices and Invariants*, pp. 81–82.

and S_{ij} is the cofactor of s_{ij} in S, then

$$(2) \qquad \sum_{i=1}^{r} S_{ij} g_{iq} = 0, \quad (j = 1, 2, \cdots, r; q = 1, 2, \cdots, n - r).$$

For a fixed q, the equation (2) represents a set of r homogeneous equations in the r unknowns g_{iq}, and since $|S_{ij}| = |S|^{r-1} \neq 0$, it follows that $g_{iq} = 0$. Accordingly $G = 0$ and by a similar argument $H = 0$, so that F has the form

$$\begin{pmatrix} S & 0 \\ 0 & K \end{pmatrix}.$$

But, since S is non-singular, at least one of the quantities $S_{ij} \neq 0$. If k is any element of K, we observe that kS_{ij} is an element of F_s which must be zero, and therefore $K = 0$.

In case (c), the sufficiency of the condition is an immediate consequence of (1). If the rank r of A is greater than s, there must exist in A a submatrix T of $s+1$ rows and columns, which is non-singular. Without any loss of generality we may suppose that

$$A = \begin{pmatrix} T & K \\ L & M \end{pmatrix}, \qquad B = \begin{pmatrix} S & H \\ P & Q \end{pmatrix},$$

where S is an $(s+1)$-rowed square matrix. From $A_s = B_s$, we deduce that $T_s = S_s$ and

$$|T_s| = |T|^s = |S_s| = |S|^s,$$

so that

$$(3) \qquad |S| = \omega |T|,$$

where ω is an sth root of unity. Moreover*

$$(T_s)_s = |T|^{s-1} T = (S_s)_s = |S|^{s-1} S,$$

so that, by (3), $S = \omega T$. Since T is non-singular, there must exist in T a non-singular submatrix T' of s rows and columns. If A' denote a matrix obtained from A by a rearrangement of rows and columns, so that T' occurs in the top left-hand corner of A', and B' is the matrix obtained from B by exactly the same rearrangement, then

* $(T_s)_s$ denotes the sth compound of T_s. That $(T_s)_s = |T|^{s-1} T$ is simply the well known theorem on the adjugate of the adjugate of a matrix.

$$A' = \begin{pmatrix} T' & K' \\ L' & M' \end{pmatrix}, \quad B' = \begin{pmatrix} \omega T' & H' \\ P' & Q' \end{pmatrix},$$

and from $A_s = B_s$ it follows that $A_s' = B_s'$. If

$$T' = (t_{ij}), \quad K' = (k_{iq}), \quad H' = (h_{iq}),$$
$$(i, j = 1, 2, \cdots, s; q = 1, 2, \cdots, n - s),$$

and T_{ij} denote the cofactor of t_{ij} in T', then

$$\sum_{i=1}^{s} T_{ij} k_{iq} = \sum_{i=1}^{s} \omega^{s-1} T_{ij} h_{iq}, \quad \text{or} \quad \sum_{i=1}^{s} T_{ij}(k_{iq} - \omega^{s-1} h_{iq}) = 0.$$

But, since $|T_{ij}| \neq 0$, $k_{iq} - \omega^{s-1} h_{iq} = 0$ or $H' = \omega K'$. Similarly it may be shown that $P' = \omega L'$. Let T'' be a submatrix of T' of order $s-1$ which is non-singular. If m_{ij} is any element of M' and q_{ij} the corresponding element of Q', the determinant of order s formed from A' of the $s-1$ rows and columns of which T'' is composed and the row and column in which m_{ij} lies is equal to the corresponding determinant formed from B'. But from the equality of these two determinants it follows that $m_{ij}|T''| = \omega^{s-1} q_{ij} |T''|$ and therefore, since $|T''| \neq 0$, it follows that $Q' = \omega M'$, $A' = \omega B'$, and $A = \omega B$. This completes the proof of the theorem.

THE JOHNS HOPKINS UNIVERSITY

REMARKS ON PROPOSITIONS *1·1 AND *3·35 OF PRINCIPIA MATHEMATICA†

BY B. A. BERNSTEIN

1. *Object.* Among the propositions of the theory of deduction underlying Whitehead and Russell's *Principia Mathematica* are the two following:

*1·1. *Anything implied by a true elementary proposition is true.*

*3·35. $\vdash : p \cdot p \supset q \cdot \supset \cdot q$.

The authors interpret *3·35 as "if p is true, and q follows from it, then q is true," and they remark that *3·35 "differs

† Presented to the Society, September 2, 1932.

from $*1\cdot1$ by the fact that it does not apply only when *p really* is true, but requires merely the *hypothesis* that *p* is true."† It is my object to make a few remarks on these interpretations of $*1\cdot1$ and $*3\cdot35$.

2. *On Proposition* $*3\cdot35$. With regard to $*3\cdot35$, the *Principia's interpretation is inadmissible.* For in this interpretation the authors read p, $p\cdot q$, $p \supset q$, respectively, as "*p* is true," "*p* is true and *q* is true," "*p* implies *q*"; and *each of these readings violates the distinction between p and* $\vdash\cdot p$ properly made *earlier in the theory of deduction* (see the *Principia*, vol. I, pp. 91–92). Or, to state the matter in another way, the symbols p, $p\cdot q$, $p \supset q$ are each *elementary* propositions, and hence *cannot* be read, as the *Principia* reads them, as propositions *about* elementary propositions. The correct readings of p, $p\cdot q$, $p \supset q$ are, respectively: "*p*," "*p* and *q*," "not-*p* or *q*." The correct reading of $*3\cdot35$ is, then:

"The proposition 'not-[*p* and (not-*p* or *q*)] or *q*' is true."‡

3. *On Proposition* $*1\cdot1$. With regard to $*1\cdot1$, I hold that *the Principia's view that* $*1\cdot1$ *can "apply only when p really is true," is not justified.* Neither the mere wording of $*1\cdot1$ nor the use of $*1\cdot1$ in the theory of deduction justifies the view that in $*1\cdot1$ "*p really* is true." The mere wording of $*1\cdot1$ seems to me to say the same thing as the proposition *X* following:

X. If *p* is true, and *p* implies *q*, then *q* is true.

And *this proposition is true even if p is false.*

As to the *use* of $*1\cdot1$ in the theory of deduction, I have pointed out elsewhere§ that though the authors say "we cannot express this principle symbolically," they employ $*1\cdot1$ as if it were written in the form *Y* following:

Y. If $\vdash\cdot p$ and $\vdash\cdot p \supset q$, then $\vdash\cdot q$.‖

---

† See the *Principia*, p. 110. (Here and in later footnotes the *Principia* referred to is vol. I, 2d ed.)

‡ Compare the remarks on $\vdash\cdot p$ and $p \supset q$ in my review of the revised edition of the *Principia*, this Bulletin, vol. 32 (1926), pp. 711–713. Compare also the remarks on $p \supset q$ and $p \equiv q$ in my article *On proposition* $*4.78$ *of Principia Mathematica*, this Bulletin, vol. 38 (1932), pp. 388–391.

§ In my *Whitehead and Russell's theory of deduction as a mathematical science*, this Bulletin, vol. 37 (1931), pp. 480–488.

‖ See, for example, the note following the proof of $*2.15$ in the *Principia*. See also p. xviii, regarding Nicod's "rule of inference."

But this proposition Y is simply proposition X with the primitive ideas of the theory of deduction symbolized.† Hence, the *use of* $*1\cdot1$, as well as the wording of $*1\cdot1$, fails to support the view that in $*1\cdot1$ "p really is true."

Observe that X is precisely the proposition which the authors of the *Principia* take as the interpretation of $*3\cdot35$. We thus find that the *Principia's* interpretation of $*3\cdot35$, an interpretation which the authors hold to be inapplicable to $*1\cdot1$, *really fits* $*1\cdot1$ *and does not fit* $*3\cdot35$.

4. *On* $*1\cdot1$ *and* $1\cdot1$. It should be noted that proposition $*1\cdot1$ is not the same as proposition $1\cdot1$, the proposition that I used previously‡ as the Boolean form of $*1\cdot1$. Proposition $1\cdot1$ is:

$1\cdot1$. There exists a K-element 1 such that from $p=1$ and $p'+q=1$ follows $q=1$.

This proposition is not a mere Boolean symbolization of $*1\cdot1$. Such a symbolization would be simply:

Z. From $p=1$ and $p'+q=1$ follows $q=1$.

That is, the mere Boolean symbolization of $*1\cdot1$ would be simply proposition $1\cdot1$ with the omission of the clause stating that the element 1 exists. This clause, however, as I explained in my paper (loc. cit.) dealing with the Boolean translation of the theory of deduction, is involved in the *Principia's* $*1\cdot1$. My proposition $1\cdot1$ is thus neither $*1\cdot1$ nor $*3\cdot35$; but it is $*1\cdot1$ in which is made explicit the fact, implied in $*1\cdot1$, that there exists a "true" proposition 1.§

† Note that $\vdash\cdot p \supset q$ may properly be read "p implies q." (For a discussion of the symbolization of "p implies q" see my review of the *Principia*, loc. cit., and my article on proposition *4.78, loc. cit.) Note also that the unsymbolized "if . . . then" and "and" are properly *outside* the theory of deduction. (For a discussion of the ideas *within* a mathematical science and the ideas *outside* the science see my *Whitehead and Russell's theory of deduction as a mathematical science*, loc. cit.)

‡ In my *Whitehead and Russell's theory of deduction as a mathematical science*, loc. cit.

§ Paul Henle (this Bulletin, vol. 38 (1932), p. 409) says of my proposition $1\cdot1$ that it "is not an accurate transcription of $*1\cdot1$ unless the convention be adopted that the postulate is not satisfied by any case in which the *hypothesis* is not satisfied," and he merely refers to p. 110 of the *Principia*, the page containing the interpretation of $*3\cdot35$ and the remark concerning the relation of $*3\cdot35$ to $*1\cdot1$ quoted in §1 above. It seems to me that the question of

5. *On* $*1\cdot1$ *and* $*1\cdot11$. It should be noted further that, unless the authors use their $*1\cdot1$ in some such form as Y above, they will not have provided proofs for a host of theorems in Section A of the *Principia*. For they use $*1\cdot11$ in the proofs of many theorems; but, while they tell us in the second edition of the *Principia* to drop $*1\cdot11$ from the list of primitive propositions, *they seemingly fail to tell us explicitly what to substitute for* $*1\cdot11$ *in the proofs employing it.* Proposition Y above seems to me well suited as the desired substitute. As an illustration of the working of Y in place of $*1\cdot11$, see my derivation† of Nicod's postulates from the primitives of the *Principia*. I may add that that paper summarizes in a way my position with regard to the mathematics of Section A of the *Principia*.‡

---

the accuracy of $1\cdot1$ as a translation of $*1\cdot1$ should not be decided by "convention," unless the language of $1\cdot1$ is ambiguous. But it is clear to me that my $1\cdot1$ is not ambiguous, and *is* "satisfied by any case in which the hypothesis is not satisfied." Nevertheless, I hold that $1\cdot1$ is an accurate translation of $*1\cdot1$, as I have tried to show in the above discussions of $*1\cdot1$ and $1\cdot1$.

† Bernstein, *On Nicod's reduction in the number of primitives of logic*, Proceedings of the Cambridge Philosophical Society, vol. 28 (1932), pp. 427–432.

‡ Since the above was written, an admirable paper by Huntington has appeared (Transactions of this Society, vol. 35 (January, 1933), pp. 274–304), in which a position is taken regarding the *Principia* somewhat different from mine. In Appendix II, Huntington presents a "set of postulates from which all the propositions, both 'formal' and 'informal,' in Section A of the *Principia*, are deducible." The postulates $1\cdot1$–$1\cdot71$ which form my version of Section A, however, I based on the *formal* propositions in Section A, in accordance with the *Principia's* statement (p. vii) that "our logical system is wholly contained in the numbered propositions, which are independent of the Introduction and the Summaries. The Introduction and the Summaries are wholly explanatory, and form no part of the chain of deductions." With regard to Huntington's observation (p. 291) that my *Principia* postulate $1\cdot5$ is derivable from my other postulates (in accordance with Bernays' findings in examining the *Principia* primitives themselves), I may say that in my paper in the June, 1931, issue of this Bulletin, the chief aim was to express the *Principia* primitives in customary mathematical language; that the redundancies removed there are only *obvious* redundancies discovered in the pursuance of the main aim of the paper; and that at the time of writing the paper, I was unaware of Bernays' work.

THE UNIVERSITY OF CALIFORNIA

ON THE CLASS NUMBERS OF A CYCLIC FIELD AND A SUB-FIELD*

C. G. LATIMER

1. *Introduction.* Let F_1 be an algebraic field which is cyclic with respect to the rational field and let F_2 be a sub-field of F_1. Kummer stated that if F_1 is a divisor of the field defined by a λth root of unity, λ a prime, then the class number of F_2 is a divisor of the class number of F_1.† He employed the broader definition of equivalence. However, as pointed out by Hilbert,‡ there is an error in his proof. Furtwängler proved the theorem for the case where λ is a power of a prime, using narrow equivalence.§

The purpose of this paper is to prove the following theorem which overlaps but does not include Furtwängler's.

THEOREM. *Let F_1 be a field which is cyclic with respect to the rational field and such that the discriminant of every sub-field, not rational, contains a prime factor not a divisor of the degree F_1. If F_2 is a sub-field of F_1 and if h_1, h_2 are the number of classes of narrowly equivalent ideals in F_1, F_2, respectively, then h_2 is a divisor of h_1.*

Furtwängler gave an example of a non-cyclic abelian field F_1, for which this theorem is not valid.‖

2. *A Lemma.* In the next paragraph, the above theorem will be proved by use of the following lemma and a theorem due to Chevally.

---

* Presented to the Society, December 27, 1932.

† Journal für Mathematik, vol. 40 (1850), pp. 114–6; Bulletin of the National Research Council, No. 62, *Algebraic Numbers*, II, Vandiver and Wahlin, p. 16.

‡ *Bericht über die Theorie der algebraischen Zahlkörper*, p. 378.

§ Journal für Mathematik, vol. 134 (1908), pp. 91–94. In this article, Furtwängler states (p. 91) that Kummer's theorem is correct since his result is a generalization of Kummer's. Since he and Kummer used different definitions of equivalence, it is not obvious that his theorem includes Kummer's and the validity of the latter is still an open question.

‖ Loc. cit., p. 94.

LEMMA. *Let K be a field such that $F_1 \geqq K > F_2$. If K is of degree e with respect to a sub-field F, where e is a prime, and if the discriminants of K and F are D and d respectively, then $D \neq d^e$.*

Every prime factor of d is a factor of D. If D contains a prime factor not a divisor of d, the lemma follows at once. We shall therefore assume hereafter that a prime divides D only if it divides d.

F_1 is cyclic and hence the same is true of K and F. Let F be of degree $E = e_1 \cdot e_2 \cdots e_m$, where the e's are powers of distinct primes. Since F is an abelian field, it may be obtained by composition of cyclic abelian fields of prime power degree and these prime powers are the invariants of the Galois group of F.* Since this group is cyclic, these invariants are relatively prime in pairs. It follows that $e_1, e_2, \cdots, e_m$ are the invariants of the group and F is obtained by composition of m cyclic fields $F^{(e_1)}$, $F^{(e_2)}, \cdots, F^{(e_m)}$ of degrees $e_1, e_2, \cdots, e_m$, respectively.

The degree eE of K is a divisor of the degree of F_1 and therefore, by the hypothesis of the theorem, D contains a prime factor p, not a divisor of eE. Then p is a divisor of d and hence it is a divisor of the discriminant of at least one of the fields $F^{(e_i)}$. After proper choice of notation, we may assume that p is a divisor of the discriminant of each of the fields $F^{(e_i)}, (i = 1, 2, \cdots, s \leqq m)$, and, if $s < m$, it is not a divisor of the discriminant of any of the fields $F^{(e_i)}$, $(i = s+1, \cdots, m)$.

Since every $F^{(e_i)}$ is a cyclic field of prime power degree, it follows that

$$p = \mathfrak{P}_i^{e_i}, \qquad (i = 1, 2, \cdots, s),$$

where $\mathfrak{P}_i$ is a prime ideal in $F^{(e_i)}$.† Employing the fact that the e_i are relatively prime in pairs, we may show that if $F^{(E_1)}$ is the field of degree $E_1 = e_1 \cdot e_2 \cdots e_s$ obtained by composition of the $F^{(e_i)}$, $(i = 1, 2, \cdots, s)$, then

$$p = \mathfrak{P}'^{E_1},$$

where $\mathfrak{P}'$ is a prime ideal in the field $F^{(E_1)}$. If $E = E_1$, d is divisible by p^{E-1} and by no higher power of p.‡ Suppose E/E_1

* Weber, *Lehrbuch der Algebra*, 2d edition, vol. 2, p. 764.

† Weber, *Zur Theorie der zyklischen Zahlkörper*, Mathematische Annalen, vol. 67 (1909), p. 48.

‡ Bachmann, *Allgemeine Arithmetik der Zahlenkörper*, p. 294.

$=E_2>1$. Let $F^{(E_2)}$ be the field obtained by composition of the fields $F^{(e_i)}$, $(i=s+1, \cdots, m)$. Since p is not a divisor of the discriminant of any of these $m-s$ fields, it is not a divisor of the discriminant of $F^{(E_2)}$. Hence

$$p = \mathfrak{P}_1'' \cdot \mathfrak{P}_2'' \cdot \cdots \mathfrak{P}_k'',$$

where the $\mathfrak{P}_i''$ are distinct prime ideals in $F^{(E_2)}$, each of degree $t=E_2/k$. The field F is obtained by composition of $F^{(E_1)}$, $F^{(E_2)}$. Employing the above factorizations of p in these fields, it may be shown that

$$p = (\mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdot \cdots \mathfrak{p}_k)^{E_1},$$

where the $\mathfrak{p}_i$ are distinct prime ideals in F, each of degree t. It follows that

$$d = p^{E_2(E_1-1)} \cdot d_1 = p^{E(1-E_1^{-1})} \cdot d_1,$$

where d_1 is prime to p.* This holds if $E=E_1$.

Consider the field K. It is of degree eE. If e is prime to E, it may be shown, as for F, that K is obtained by composition of the above fields $F^{(e_i)}$, and another cyclic field $F^{(e)}$ of degree e. If e is not prime to E, it divides one of the e_i, say e_1. Then it may be shown that K is obtained by composition of a cyclic field $F^{(ee_1)}$ of degree ee_1, and the fields $F^{(e_2)}$, $F^{(e_3)}$, $\cdots$, $F^{(e_m)}$.

We have assumed before that p is a divisor of D. By the condition of the theorem on the sub-fields of F_1, we may assume that it is also a divisor of the discriminant of $F^{(e)}$, or $F^{(ee_1)}$. We find, as for F, that

$$D = p^{eE(1-E_1'^{-1})} \cdot D_1,$$

where D_1 is prime to p and E_1' is the product of the degrees of those fields $F^{(e)}$, $F^{(e_1)}$, $\cdots$, $F^{(e_m)}$ or $F^{(ee_1)}$, $F^{(e_2)}$, $\cdots$, $F^{(e_m)}$ whose discriminants are divisible by p. Since one of these fields is $F^{(e)}$, or $F^{(ee_1)}$, it follows that $E_1' \neq E_1$. The lemma follows from the above expressions for d and D.

3. *Proof of Theorem.* Let K be a field such that $F_1 \geqq K > F_2$. There is a sequence of fields $K=K_0>K_1> \cdots >K_t=F_2$ such that the degree of every K_{i-1} with respect to K_i is a prime $e^{(i)}$. If D_j is the discriminant of K_j, then

$$D_{i-1} = D_i^{e^{(i)}} N(\mathfrak{D}_i), \qquad (i = 1, 2, \cdots, t),$$

* Bachmann, loc. cit.

where $\mathfrak{D}_i$ is* the relative discriminant of K_{i-1} with respect to K_i. The field K is of degree $\epsilon = e^{(1)} \cdot e^{(2)} \cdots e^{(t)}$ with respect to F_2 and by the last reference

$$D = d^{\epsilon} N(\mathfrak{D}),$$

where $\mathfrak{D}$ is the relative discriminant of K with respect to F_2. By the Lemma, every $N(\mathfrak{D}_i) > 1$. It follows that $N(\mathfrak{D}) > 1$. But, by a result due to Chevally, h_2 divides h_1 if there is no field K, $F_1 \geqq K > F_1$, such that the relative discriminant of K with respect to F_2 is of norm unity.† The theorem follows.

UNIVERSITY OF KENTUCKY

SOME APPLICATIONS OF MURPHY'S THEOREM‡

BY H. BATEMAN

It is well known that the linear partial differential equations of mathematical physics possess solutions in the form of definite integrals with limits depending on the variables entering into the partial differential equations. The law connecting the limits of such an integral with the integrand looks at first sight rather mysterious but the whole matter becomes clear when the integral is expressed as a contour integral with the aid of a theorem due to Murphy and, in a slightly different form, to Cauchy.§

If C is a closed contour containing just one root, a, of the equation $F(x) = 0$ and just one root, b, of the equation $G(x) = 0$, then, if the radii from these roots turn completely round just once and in one direction as a point describes this contour and if the functions $f(z)$, $\int f(z)dz$, $F(z)$, and $G(z)$ are analytic and uni-

---

* Bachmann, loc. cit., p. 452.

† Chevally, *Relation entre le nombre de classes d'un sous-corps et celui d'un sur-corps*, Comptes Rendus, vol. 192 (1931), pp. 257–258.

‡ Presented to the Society, December 27, 1932.

§ R. Murphy, Transactions of the Cambridge Philosophical Society, vol. 3 (1830), p. 429, A. L. Cauchy, Journal de l'École Polytechnique, vol. 12 (1823), p. 580. Murphy's integral has been transformed into a contour integral from which Cauchy's relation may be obtained by an integration by parts.

form at all points inside C and on this contour, we have

$$(1) \qquad \int_a^b f(x)dx = \frac{1}{2\pi i}\int_C f(z)dz \log\frac{G(z)}{F(z)}.$$

To apply this theorem to Laplace's equation in n variables,

$$(2) \qquad \frac{\partial^2 V}{\partial x_1^2} + \frac{\partial^2 V}{\partial x_2^2} + \cdots + \frac{\partial^2 V}{\partial x_n^2} = 0,$$

we make use of the fact that if $c_1, c_2, \cdots, c_n$ are constants satisfying the equation $c_1^2+c_2^2+\cdots+c_n^2=0$, the function

$$V = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1-n/2} \cdot \log\frac{(x_1-c_1)^2+(x_2-c_2)^2+\cdots+(x_n-c_n)^2}{x_1^2 + x_2^2 + \cdots + x_n^2}$$

is a particular solution. This is easily verified by noticing that if $r^2=x_1^2+x_2^2+\cdots+x_n^2$, then, for sufficiently large values of r, the function V can be expanded in a power series of type

$$\sum A_p(c_1x_1 + c_2x_2 + \cdots + c_nx_n)^p r^{2-n-2p},$$

and each term of this series is a solution of the equation.

Generalizing this solution we consider the integral

$$V = \frac{1}{2\pi i}\int_C R^{2-n}f(z)dz\left[\log\frac{H^2}{R^2} - \log\frac{K^2}{R^2}\right],$$

where

$$R^2 = \sum_{m=1}^{n}[x_m - \xi_m(z)]^2, \quad H^2 = \sum_{m=1}^{n}[x_m - c_m(s)]^2,$$

$$K^2 = \sum_{m=1}^{n}[x_m - e_m(t)]^2,$$

and $\xi_m(z)$, $c_m(s)$, $e_m(t)$ are analytic functions of their arguments in suitable domains. The quantities s and t are, moreover, given as functions of z by the equations

$$\sum_{m=1}^{n}[c_m(s) - \xi_m(z)]^2 = 0, \quad \sum_{m=1}^{n}[e_m(t) - \xi_m(z)]^2 = 0.$$

The contour C is chosen so that $f(z)R^{2-n}$ is analytic in the en-

closed realm, so that the equation $R=0$ has no root within it, and so that $\log (H^2/K^2)$ is an analytic function of z which is infinite only at two points within C and returns to its initial value when a point goes round C. We may also consider an integral

$$V = \frac{1}{2\pi i}\int_C R^{2-n} f(z)dz \log \frac{U}{W},$$

where

$$U = \sum_{m=1}^{n} l_m(z)[x_m - u_m(z)], \quad W = \sum_{m=1}^{n} p_m(z)[x_m - v_m(z)],$$

$$\sum_{m=1}^{n} [l_m(z)]^2 = 0, \quad \sum_{m=1}^{n} [p_m(z)]^2 = 0.$$

In both cases the integrand of the integral V is a solution of (2) for all values of z not depending on the variables x.

The results obtained by considering these two integrals imply that the integral

$$V = \int_0^a \frac{f(z)dz}{R^{n-2}}$$

is a solution of equation (2) if a is defined either by the equations

$$\sum l_m(a)[x_m - u_m(a)] = 0, \quad \sum [l_m(a)]^2 = 0,$$

or by the equations

$$\sum [x_m - c_m(s)]^2 = 0, \quad \sum [c_m(s) - \xi_m(a)]^2 = 0.$$

These results are well known*; they are generally more useful when applied to the equation of wave motion than when applied to Laplace's equation, because in the former case the limits of the integral may be real.

In the case of the equation

$$F\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)V = 0, \tag{3}$$

where $F(x, y, z)$ is a homogeneous polynomial of degree n in x, y, and z, we may form a solution of type

* H. Bateman, this Bulletin, vol. 24 (1918), p. 296; The Tôhoku Mathematical Journal, vol. 13.

$$V = \frac{1}{2\pi i}\int_C ds f[x\xi(s) + y\eta(s) + z\zeta(s), s] \log \frac{x\xi(s) + y\eta(s) + z\zeta(s) - \phi(s)}{x\xi(s) + y\eta(s) + z\zeta(s) - \psi(s)},$$

where

$$F[\xi(s), \eta(s), \zeta(s)] \equiv 0.$$

An evaluation of the integral by means of Murphy's theorem gives the solution

$$V = \int_a^b f[x\xi(s) + y\eta(s) + z\zeta(s), s]ds,$$

where the limits a and b are defined by the equations

$$x\xi(a) + y\eta(a) + z\zeta(a) - \phi(a) = 0,$$
$$x\xi(b) + y\eta(b) + z\zeta(b) - \psi(b) = 0,$$

respectively. This result has already been established in another way with the aid of an extension of Lagrange's expansion.* Murphy's theorem fails when the limits of the integral are roots of the same equation $F(x) = 0$. In this case there is a slight modification of the formula in which a double loop integral is used. If the contour C is replaced by a contour shaped like a figure of eight, with a root of $F(x) = 0$ in each oval, the integral

$$\frac{1}{2\pi i}\int_\infty f(z)dz \log F(z) = \int_a^b f(x)dx,$$

where a and b are roots of $F(x) = 0$ contained in the ovals of 8. This equation may be derived from Cauchy's equation by an integration by parts.

Murphy's formula needs modification also when the function $f(z)$ is not uniform. The type of modification which is needed will be shown by a consideration of the following example. Let

$$I = \int_{-a}^{a} \frac{\phi(x)dx}{(a^2 - x^2)^{1/2}}.$$

* H. Bateman, Transactions of this Society, vol. 28 (1926), p. 346. The special case in which f is independent of its first argument is given by A. L. Dixon, Messenger of Mathematics, vol. 33 (1904), p. 172.

The first step is to form a uniform integrand by making the substitution $x=a\cos\theta$. We may then write

$$I=\int_0^{\pi}\phi(a\cos\theta)d\theta=\frac{1}{2\pi i}\int_C\phi(a\cos\tau)\log\left[\frac{\cos(\frac{1}{2}\tau)}{\sin(\frac{1}{2}\tau)}\right]d\tau,$$

where C is a simple contour enclosing the points $\tau=0$ and $\tau=\pi$ but no point of type $\tau=n\pi$ where n is different from 0 and 1. Transforming back to the variable z given by $z=a\cos\tau$, we obtain

$$I=\frac{1}{4\pi i}\int_\Gamma\frac{\phi(z)}{(a^2-z^2)^{1/2}}\log\left(\frac{a+z}{a-z}\right)dz,$$

where Γ is a contour which crosses itself at two points A and B so as to form two enclosed loops one of which contains the point $-a$ and the other the point a. By using a similar type of contour with a loop enclosing the branch point, we may write with $n=3$

$$V=\int_a^b\frac{f(z)dz}{R}=\frac{1}{4\pi i}\int_\Gamma\frac{f(z)dz}{R}\log\frac{H^2}{R^2},$$

where b is a value of z which makes $R^2=0$ and a is a value of z which makes $H^2=0$. Since the integral with limits 0 and a has been shown to be a solution of Laplace's equation, it follows that the integral

$$V=\int_0^a\frac{f(z)dz}{R}$$

is a solution of Laplace's equation. Particular cases of this theorem have been known for some time, but a simple proof of the general theorem has been lacking. A long proof by direct differentiation may be obtained by integrating by parts before differentiating whenever the integrand of an integral becomes infinite at b.

Murphy's formula, in the form in which it was given originally, is a simple generalization of a formula given by Parseval,*

$$y=a-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left[1-e^{iz}f(a+e^{-ix})\right]e^{-iz}dx,$$

* M. A. Parseval, Mémoires des Savants Étrangers de l'Institut de France, vol. 1 (1805), p. 567.

for the root of $y=a+f(y)$ which is expressed by means of Lagrange's expansion. This formula was extended by Poisson,* who showed that

$$-\frac{1}{2\pi}\int_{-\pi}^{\pi} F'(a+e^{-ix})\log\left[1-e^{ix}f(a+e^{-ix})\right]e^{-ix}dx$$

$$=F'(a)f(a)+\frac{1}{2!}\frac{d}{da}[F'(a)\{f(a)\}^2]+\cdots=F(y)-F(a).$$

Poisson also mentions that Cauchy had presented to the Academy of Sciences a memoir on the expression by means of definite integrals of the roots of equations of any degree. This was probably the unpublished memoir of 1819 mentioned by E. Lindelöf in his *Calcul des Résidus* (p. 21). There is, however, a memoir with precisely the above title, that was read in 1824 and described briefly in Cauchy's collected works.

Murphy's rule, that if x_1 is a root of $\phi(x)=0$, $\int_0^{x_1} f(x)dx$ is equal to the coefficient of $1/x$ in

$$-f(x)\log\left[\frac{\phi(x)}{x}\right],$$

has been discussed by Whipple,† who also expresses Murphy's theorem by a contour integral in practically the form in which it is used here except that his contour is a particular one.

CALIFORNIA INSTITUTE OF TECHNOLOGY

* S. D. Poisson, Journal de l'École Polytechnique, vol. 12 (1823), p. 497.

† F. J. W. Whipple, Quarterly Journal of Pure and Applied Mathematics, vol. 40 (1909), p. 368.

ON SINGULAR CHAINS AND CYCLES

BY S. LEFSCHETZ

1. *Introduction.* The theory of the topological invariance of the absolute or relative combinatorial characters of a complex, as developed in our Colloquium Lectures on *Topology* (Chapter II), was based, following Alexander and Veblen, upon the concept of singular chain. Our presentation, and indeed any known to us, appears to give rise to many misconceptions which it is proposed to clear up in the present note. Unless otherwise stated the notations are those of *Topology*.

2. *Singular Cells.* Let $\mathfrak{R}$ be a topological space and let e_p be a simplicial oriented cell such that there exists a continuous single-valued transformation (=c.s.v.t.) T of the point set e_p into a subset E_p of $\mathfrak{R}$, where $E_p = Te_p$. The symbol (e_p, T, E_p), associated with the set $\overline{E}_p$ is called a *singular oriented p-cell on* $\mathfrak{R}$. If e_p' is another e_p, there exists a barycentric transformation U of $\bar{e}_p'$ into $\bar{e}_p$: $U\bar{e}_p' = \bar{e}_p$. If we set $T' = TU$, it is evident that (e_p', T', E_p) defines also a singular oriented p-cell on $\mathfrak{R}$. We shall agree to consider it as identical with the first:

$$(1) \qquad (e_p', T', E_p) = (e_p, T, E_p).$$

This has the advantage of freeing the notion of singular cell from a too narrow connection with a specific image e_p.

3. *Singular Chains.* The singular p-chain C_p on $\mathfrak{R}$ is now defined as the association of a symbol

$$(2) \qquad C_p = \sum t_i(e_p^i, T^i, E_p^i)$$

with coefficients t belonging to one of the three rings (rational numbers, integers, integers mod m) considered in *Topology*, together with the set of all sets $\overline{E}_p^i$ corresponding to t's $\neq 0$. As a special case the e's might be cells of a finite complex k such that there exists a c.s.v.t. T of k into a subset of $\mathfrak{R}$. Then the chain symbol may take the form

$$(3) \qquad C_p = \sum t_i(e_p^i, T, E_p^i),$$

and C_p may be considered as the image of the subchain

$$c_p = \sum t_i e_p^i \tag{4}$$

of k, but that is not essential. In this instance we might have represented C_p by the symbol $(\sum t_i e_p^i, T, \sum t_i E_p^i)$ analogous to the cell symbol. Observe also that we may find for any chain (2) an equivalent representation (3). For we may take the cells e_p^i to be simplexes in some S_n whose closures do not meet, then define T as coincident with T^i on e_p^i. The closure of the sum of the cells e_p^i will then be k, and (2) will assume the form (3).

If we have several singular chains C_p^i, then $\sum s_i C_p^i$, where the coefficients s_i belong to the same ring as those of the chains, defines a p-chain which is called the linear combination with coefficients s_i of the chains C_p^i. We have thus moduli of singular p-chains wholly analogous to the moduli of subchains of a complex.

4. *Boundary Relations.* Returning to (e_p, T, E_p), let the boundary relations for e_p be

$$e_p \rightarrow \sum \eta_i e_{p-1}^i = F(e_p). \tag{5}$$

Since T is a transformation of $\bar{e}_p$ into $\overline{E}_p$, it transforms e_{p-1}^i into a subset E_{p-1}^i of $\overline{E}_p$ and hence $(e_{p-1}^i, T, E_{p-1}^i)$ is a singular $(p-1)$-cell on $\mathfrak{R}$. The singular $(p-1)$-chain

$$F(e_p, T, E_p) = \sum \eta_i (e_p^i, T, E_{p-1}^i) \tag{6}$$

is called the boundary of (e_p, T, E_p) and we write here also

$$(e_p, T, E_p) \rightarrow F(e_p, T, E_p). \tag{7}$$

The boundary of the chain (2) is now by definition

$$F(C_p) = \sum t_i F(e_p^i, T^i, E_p^i), \tag{8}$$

for which we also write

$$C_p \rightarrow F(C_p). \tag{9}$$

Owing to (1) this boundary depends solely on C_p but not on the particular transformations T^i that occur in (8). Let C_p be in the special form (3) with an associated non-singular image (4), and let the boundary relation for c_p be

$$c_p \rightarrow \sum s_i e_{p-1}^i. \tag{10}$$

Then we have well defined singular cells $(e_{p-1}^i, T, E_{p-1}^i)$ and we find that

$$C_p \rightarrow \sum s_i(e_{p-1}^i, T, E_{p-1}^i), \tag{11}$$

which may be described by the statement: the boundary of a singular image of a chain is the singular image of the boundary of the chain.

5. *Degenerate Case.* Let $\mathfrak{R}$ undergo a c.s.v.t. U into a new space $\mathfrak{R}'$. Then the singular cell (e_p, T, E_p) will go over into a singular p-cell (e_p, UT, UE_p) and C_p into

$$UC_p = \sum t_i(e_p^i, UT^i, UE_p^i). \tag{12}$$

To different representations of the same singular p-cell on $\mathfrak{R}$ there will merely correspond different representations of the same singular p-cell on $\mathfrak{R}'$, and to $F(C_p)$ there will now correspond $F(UC_p)$. In particular also if $C_p \equiv 0$, likewise $UC_p \equiv 0$.

The preceding observations have an immediate application to degenerate cells. Let (e_p, T, E_p) be a singular cell on $\mathfrak{R}$, and let us suppose that there exists a simplex σ_q, $q < p$, and two c.s.v.t.'s T', T'', such that T' is a simplicial transformation of e_p into σ_q and that $T'' \cdot T' = T$. The cell (e_p, T, E_p) is called a *singular degenerate p-cell* on $\mathfrak{R}$ and chains made up exclusively of such cells are called *degenerate chains*. If $\mathfrak{R}' = U\mathfrak{R}$ as above, the degenerate cells and chains of $\mathfrak{R}$ go over into degenerate cells and chains of $\mathfrak{R}'$.

According to *Topology*, Chapter II, No. 2, $F(e_p, T', \sigma_q)$ is a degenerate $(p-1)$-chain, and hence when (e_p, T, E_p) is degenerate so is its boundary. Hence this holds likewise as regards degenerate p-chains. Let us agree to consider all degenerate chains as identically zero. By the observation just made degenerate chains will then completely disappear from all boundary relations.

6. *Homologies.* From the preceding section it appears clearly that when $\mathfrak{R}$ and $\mathfrak{R}'$ are homeomorphic, the homeomorphism between them associates respectively to one another their moduli of p-chains, of bounding p-chains and their degenerate p-chains. These are therefore topological and the homology characters derived from the moduli are topological invariants.

Regarding these homologies, we introduce them exactly as for complexes. In particular if A intersects B in a set $A \cdot B$ closed

relatively to B, the neglect of the singular cells (e_p, T, E_p), such that $E_p \subset A$, leads to the characters of B mod A.

7. *Invariance of the Combinatorial Homology Characters.* Suppose now that $\mathfrak{R}$ itself is a finite simplicial complex K and let ϵ_p^i designate its cells. They can be considered as the singular cells $(\epsilon_p^i, 1, \epsilon_p^i)$ and it is readily seen that the formal singular boundary relations involving these cells alone are the same as the combinatorial relations between the cells ϵ_p^i themselves. Therefore whenever only singular cells of this type are involved, the singular boundary relations (7) for K are reduced to the combinatorial relations.

The invariance of the combinatorial homology characters of K is established by identifying them with the corresponding topological characters. The steps in the proof are as follows.

(a) Let C_p be a singular chain on K which we assume henceforth in the simplified form (11) with T and the non-singular prototype c_p fixed. There exists an $\eta > 0$ depending on K but not on C_p, such that when mesh $C_p < \eta$, the chain can be homotopically deformed into a subchain C_p' of K, the deformation keeping each cell on the closure of the cell of K that carries it. This is the deformation theorem (*Topology*, p. 86). It implies (loc. cit., p. 78) that there are deformation chains, all singular, indicated by $\mathfrak{D}$, such that

$$\begin{aligned} &\mathfrak{D}C_p \to C_p' - C_p - \mathfrak{D}F(C_p), \\ &C_p' = \textstyle\sum s_i E_p^i. \end{aligned} \tag{13}$$

(b) If mesh $C_p > \eta$ the chain has a subdivision chain C_p whose mesh is suitable. Subdivision is defined as in *Topology* (p. 85), by reference to a subdivision of c_p.

(c) Suppose that C_p possesses certain cells (not necessarily p-cells) which belong to K and whose sum is therefore a subcomplex K_q of K. Then the subdivision and deformation in (b) may be so chosen as to leave K_q fixed point for point. The proof indicated in *Topology* (p. 87, Remark I) only shows that C_p may be so modified as to leave the cells of K_q invariant individually but not point for point. The more accurate result, which is of interest for its own sake, is proved as follows. We show by induction as in *Topology* (p. 86) that the deformations there indicated leave K_q invariant point for point provided that

any cell of C_p without vertices on K_q is of diameter $<\eta$, and that a cell having a face in common with K_q has all its points not farther than η from that face. If C_p does not fulfill these conditions we find by reference to c_p that a suitable subdivision of C_p without new vertices on K_q will behave as required. For if c_p is a subchain, say of k, there is a subcomplex k' of k such that $T \cdot k' = K_q$. We can then apply to k a series of subdivisions differing from regular subdivisions only in so far that no new vertices are ever introduced on k'. Given any ζ we can thus obtain a subdivision c_p^* of c_p whose cells fulfill relatively to k' and ζ the two conditions that we wish to impose upon the cells of C_p relatively to K_q and η. Since T is continuous the required result follows for C_p.

Consider now the boundary relations mod L, where L is a subcomplex of K. Let Γ_p be a (singular) cycle mod L. By (a) there is a subdivision Γ_p' of Γ_p homotopically deformable into a subcycle Γ_p^* of K, its points on L remaining on L, and $\Gamma_p' \sim \Gamma_p$ mod L on Γ_p itself (*Topology*, p. 87) and hence a fortiori on K. By (c), if $C_{p+1} \to \Gamma_p'$ mod L, there is a subdivision $C_p'{}_{+1}$ of C_{p+1} with the same boundary Γ_p', deformable into a subchain $C_p^*{}_{+1}$ by a homotopy leaving Γ_p' invariant, so that $C_p^*{}_{+1} \to \Gamma_p'$ mod L. Therefore if the initial cycle ~ 0 mod L in the topological sense, the reduced cycle ~ 0 mod L in the combinatorial sense. From this follows immediately as in *Topology* (p. 88), that the topological and combinatorial homology groups of the same types are simply isomorphic and hence have the same numerical invariants. Therefore the combinatorial homology characters are topological invariants.

8. *Remarks.* I. Once the notion of singular cell has become familiar one will naturally abandon the explicit (too explicit) (e, T, E) notation in favor of the simpler E of *Topology*.

II. The following circumstance may arise in connection with our definition of singular cell. Taking for the sake of simplicity $p=2$, let $e_2 = ABC$ be an (oriented) isosceles triangle with $AB = AC$ and let AD be the altitude issued from A. Let U be the symmetry about AD and T a transformation $=1$ on ADB, $= U$ on ADC. If we set $T' = T \cdot U$, $e_2' = ACB$, we have

$$(e_2, T, E_2) = (e_2', TU, E_2) = (-e_2, TU, E_2) = (-e_2, T, E_2)$$

and therefore

$$2(e_2, T, E_2) = 0.$$

Owing to this, E. Čech, who pointed out this circumstance to us, suggested that in the present and in the similar instance for any p, the singular cell be also considered as degenerate. The more extended meaning to be thus attached to degenerate cells, while justifiable, is not however essential.

PRINCETON UNIVERSITY

VARIABLES CORRELATED IN SEQUENCE*

BY A. T. CRAIG

1. *Introduction.* If each of n variables, $x_1, x_2, \cdots, x_n$, represents a quantitative character of an individual, and if the variables are correlated in sequence, that is, x_1 is correlated with x_2, x_2 is correlated with x_3, $\cdots$, and in general x_i is correlated with x_{i+1}, it seems natural to inquire about the correlation between a character, say x_1, of one individual and a character, say x_3, of a second individual, with the condition imposed that the two individuals have identical measurements with regard to the character x_2. It is this problem with which we shall be primarily concerned in the present paper. As we proceed, we shall place appropriate restrictions upon the nature of the correlation which exists between the variables. We shall, however, make no assumptions regarding the correlation between the variables other than that between them in adjacent pairs.

In order to provide a convenient point of departure and to exhibit a set of variables correlated in sequence, we shall first consider a rather elementary problem which arises when measurements are made under a constant law of probability.

2. *The Correlation between Measurements under a Constant Law of Probability.* Let the variable t obey a constant law of probability $f(t) = 1/a$, $0 \leqq t \leqq a$. Let successive sets of n independent measurements each, say $t_1, t_2, \cdots, t_n$, be made upon t. We may, without loss of generality, suppose the meas-

* Presented to the Society, April 8, 1932, under the title *Some properties of correlated variables.*

urements of the sets to be arranged in ascending order of magnitude so that $0 \leqq t_1 \leqq t_2 \leqq \cdots \leqq t_n \leqq a$. If we denote t_p, t_q, t_r, $p \leqq q \leqq r$, by x, y, z respectively, then the simultaneous law of probability of the pth, qth, and rth measurements is given by

$$\begin{aligned}\phi(x, y, z) &= \binom{n}{p-1}\binom{n-p+1}{q-p-1}\binom{n-q+2}{r-q-1}\binom{n-r+3}{n-r}\binom{3}{1}\binom{2}{1} \\ &\cdot\left[\int_0^x f(t)dt\right]^{p-1}\left[\int_z^y f(t)dt\right]^{q-p-1}\left[\int_y^z f(t)dt\right]^{r-q-1} \\ &\cdot\left[\int_z^a f(t)dt\right]^{n-r} f(x)f(y)f(z) \\ &= Kx^{p-1}(y-x)^{q-p-1}(z-y)^{r-q-1}(a-z)^{n-r},\end{aligned}$$

where

$$K = \frac{n!}{(p-1)!(q-p-1)!(r-q-1)!(n-r)!a^n}.$$

The simultaneous laws of probability of the variables taken in pairs are then

$$\begin{aligned}\phi_1(x, y) &= \int_y^a \phi(x, y, z)dz = K_1x^{p-1}(y-x)^{q-p-1}(a-y)^{n-q}, \\ \phi_2(x, z) &= K_2x^{p-1}(z-x)^{r-p-1}(a-z)^{n-r}, \\ \phi_3(y, z) &= K_3y^{q-1}(z-y)^{r-q-1}(a-z)^{n-r},\end{aligned}$$

where

$$K_1 = \frac{n!}{(p-1)!(n-q)!(q-p-1)!a^n},$$

$$K_2 = \frac{n!}{(p-1)!(n-r)!(r-p-1)!a^n},$$

$$K_3 = \frac{n!}{(q-1)!(n-r)!(r-q-1)!a^n}.$$

Further, the laws of probability of x, y, and z are respectively

$$\begin{aligned}\phi_4(x) &= K_4x^{p-1}(a-x)^{n-p}, \\ \phi_5(y) &= K_5y^{q-1}(a-y)^{n-q}, \\ \phi_6(z) &= K_6z^{r-1}(a-z)^{n-r},\end{aligned}$$

where

$$K_4 = \frac{n!}{(p-1)!(n-p)!a^n},$$

$$K_5 = \frac{n!}{(q-1)!(n-q)!a^n},$$

$$K_6 = \frac{n!}{(r-1)!(n-r)!a^n}.$$

For an assigned value of y, the mean of the array of x is

$$\bar{x}_y = \frac{\int_0^y x\phi_1(x, y)dx}{\phi_5(y)} = \frac{py}{q}.$$

Similarly,

$$\bar{y}_z = \frac{(n-q+1)x + a(q-p)}{n-p+1}.$$

Thus of the two variables x and y, the regression of each on the other is linear and the coefficient of correlation is*

$$r_{xy} = \left[\frac{p(n-q+1)}{q(n-p+1)}\right]^{1/2}.$$

By the same procedure, we find

$$r_{yz} = \left[\frac{q(n-r+1)}{r(n-q+1)}\right]^{1/2}, \quad r_{xz} = \left[\frac{p(n-r+1)}{r(n-p+1)}\right]^{1/2}.$$

It may be observed that

$$r_{xz} = r_{xy}r_{yz}.$$

We shall next determine the regression surfaces of each variable on the other two. For assigned values of y and z, the mean of the array of x is

$$\bar{x}_{yz} = \frac{\int_0^y x\phi(x, y, z)dx}{\phi_3(y, z)} = \frac{py}{q}.$$

* See Karl Pearson, *On the mean character and variance of a ranked individual*, etc., Biometrika, vol. 23 (1931), p. 391.

Similarly

$$\bar{y}_{xz} = \frac{z(q-p) + x(r-q)}{r-p},$$

and

$$\bar{z}_{xy} = \frac{y(n-r+1) + a(r-q)}{n-q+1}.$$

Thus the regression surfaces are planes and consequently the multiple and partial correlation coefficients may be computed. Indeed $r_{x\cdot yz} = r_{xy\cdot z} = r_{xy}$.

If we set $u = y - x^*$ and $v = z - y$, then $\psi(u, v)$, the simultaneous probability function of the differences between the qth and pth measurements and the rth and qth measurements, is readily found. We have

$$\psi_1(u, v, z) = \phi(z - u - v, z - v, z), \quad u + v \leqq z \leqq a.$$

Hence,

$$\psi(u, v) = \int_{u+v}^{a} \phi(z - u - v, z - v, z)dz$$

$$= \frac{n!\,u^{q-p-1}(a - u - v)^{n+p-r}}{(q-p-1)!(r-q-1)!(n+p-r)!a^n}.$$

We observe that the surface $w = \psi(u, v)$ is limited by the lines $v = 0$, $u = 0$, $u + v = a$. It follows immediately that the regression of each variable on the other is linear and that

$$r_{uv} = -\left[\frac{(q-p)(r-q)}{(n+p+1-q)(n+q+1-r)}\right]^{1/2}.$$

As a special application of the function $\phi(x, y, z)$ to the statistical theory of sampling, we may take

---

* The determination of the mean value of the difference between the pth and the $(p+1)$st items is known as the Galton difference problem. This problem has been extensively studied and we cite the following references:

Karl Pearson, *Note on Francis Galton's problem*, Biometrika, vol. 1 (1901–1902), pp. 390–399; *On the probable errors of frequency constants*, Biometrika, vol. 13 (1920), pp. 113–132; *On the mean character and variance of a ranked individual*, Biometrika, vol. 23 (1931), pp. 364–397; vol. 24 (1932), pp. 203–297.

H. L. Rietz, *On a mean difference problem that occurs in statistics*, American Mathematical Monthly, vol. 17 (1910), pp. 235–240.

T. Hojo, *Distributions of medians, quartiles and interquartile distance*, Biometrika, vol. 23 (1931), pp. 315–360.

$$n = 2m + 1, \qquad p = 1, \qquad q = m + 1, \qquad r = 2m + 1$$

and determine the correlation function $F(\xi, W)$ of the range $W = z - x$ and the median $\xi = y$ in samples of $2m+1$ items drawn at random from a universe characterized by a constant law of probability. We then have

$$F(\xi, W) = \int \phi(x, \xi, x + W)dx.$$

The limits of integration are given below for ξ and W in that part of the ξW-plane bounded by the lines indicated at the right:

$\xi - W, \xi$	$W = 0, \quad W = \xi, \quad W = a - \xi;$
$0, \xi$	$\xi = 0, \quad W = \xi, \quad W = a - \xi;$
$0, a - W$	$W = a, \quad W = \xi, \quad W = a - \xi;$
$\xi - W, a - W$	$\xi = a, \quad W = \xi, \quad W = a - \xi.$

Let us now consider the problem of determining the correlation between the pth measurement of one set and the rth measurement of a second set with the condition that the two sets have identical qth measurements. Since x and z are to be matched as to y, we first choose at random a value of y. The probability that y lies in the interval $(y, y+dy)$ is, to within infinitesimals of higher order, $\phi_5(y)dy$. From the array of x corresponding to this value of y, we choose at random a value of x. The probability that x lies in the interval $(x, x+dx)$ is, to a first approximation, $[\phi_1(x, y)/\phi_5(y)]dx$. Finally, from the array of z corresponding to the same value of y, we choose at random a value of z. The probability that z lies in the interval $(z, z+dz)$ is, to a first approximation, $[\phi_3(y, z)/\phi_5(y)]dz$. The probability of the joint occurrence of these events is then the product of the separate probabilities and we have for the simultaneous probability function of x and z when chosen in this manner

$$\bar{\phi}_2(x, z) = \int_z^z \frac{1}{\phi_5(y)} \phi_1(x, y)\phi_3(y, z)dy.$$

Upon performing the integration, we observe that

$$\bar{\phi}_2(x, z) = \phi_2(x, z).$$

Accordingly, if $\bar{r}_{xz}$ be the coefficient of correlation between x and z when they are matched identically with respect to y, then

$$\bar{r}_{xz} = r_{xz} = r_{xy}r_{yz}.$$

If then, we regard x, y, z as three variables correlated in sequence with no information concerning the correlation between x and z, we may say that the coefficient of correlation between the x of one set and the z of another set, the two sets having identical y's, is equal to the product of the coefficients of correlation between x and y and between y and z.

3. *Variables Normally Correlated in Sequence.* Let

$$x_1, x_2, \cdots, x_n$$

be n variables normally correlated in sequence. That is, there are given $n-1$ normal correlation functions

$$F_i(x_i, x_{i+1}) = \frac{1}{2\pi\sigma_i\sigma_{i+1}(1 - r_i^2)^{1/2}} e^{\lambda}, \quad (i = 1, 2, \cdots, n - 1),$$

where

$$\lambda = -\frac{1}{2(1 - r_i^2)}\left\{\frac{x_i^2}{\sigma_i^2} + \frac{x_{i+1}^2}{\sigma_{i+1}^2} - \frac{2r_i x_i x_{i+1}}{\sigma_i\sigma_{i+1}}\right\},$$

and where r_i is the coefficient of correlation between x_i and x_{i+1}. If it is desired, $x_1, x_2, \cdots, x_n$ may be regarded as the pth, qth, rth, $\cdots$, sth measurements in successive sets of N. We do not, however, assume a knowledge of probability functions of more than two variables. For example, we have no function comparable to the function $\phi(x, y, z)$ of §2. Let it be required to determine the correlation between x_{j-1} of one set of measurements and x_{j+1} of another set, the two sets having identical values for x_j. As will appear obvious presently, it is sufficient to consider the variables x_1, x_2, x_3. We first choose at random a value of x_2. For this assigned x_2, we choose at random a value of x_1 and a value of x_3 from the arrays corresponding to the assigned x_2. Then, to a first approximation, the probability that x_2 lies in the interval (x_2, x_2+dx_2), that x_1 lies in the interval (x_1, x_1+dx_1), and that x_3 lies in the interval (x_3, x_3+dx_3) is

$$f_2(x_2)dx_2 \frac{1}{f_2(x_2)} F_1(x_1, x_2)dx_1 \frac{1}{f_2(x_2)} F_2(x_2, x_3)dx_3,$$

where

$$f_2(x_2) = \int_{-\infty}^{\infty} F_1(x_1, x_2)dx_1 = \int_{-\infty}^{\infty} F_2(x_2, x_3)dx_3.$$

Accordingly, the correlation function $\psi_{13}(x_1, x_3)$ of x_1 and x_3 when matched in this manner is given by

$$\psi_{13}(x_1, x_3) = \int_{-\infty}^{\infty} \frac{1}{f_2(x_2)} F_1(x_1, x_2)F_2(x_2, x_3)dx_2$$

$$= \frac{1}{2\pi\sigma_1\sigma_3(1 - r_1^2r_2^2)^{1/2}} e^{\mu},$$

where

$$\mu = -\frac{1}{2(1 - r_1^2r_2^2)}\left\{\frac{x_1^2}{\sigma_1^2} + \frac{x_3^2}{\sigma_3^2} - \frac{2r_1r_2x_1x_3}{\sigma_1\sigma_3}\right\}.$$

Thus the variables are normally correlated with the coefficient of correlation $r = r_1r_2$. If then, the variables x_{j-1} and x_{j+1} are matched with respect to x_j, the coefficient of correlation is $r = r_{j-1}r_j$.

We may now, by making use of $\psi_{13}(x_1, x_3)$ and $F_3(x_3, x_4)$, match in a somewhat analogous manner x_1 and x_4 identically as to x_3 and obtain $\psi_{14}(x_1, x_4)$ which is a normal correlation function with $r = r_1r_2r_3$. Clearly the procedure may be continued so as to involve any two of the n variables.

4. *Variables Non-Normally Correlated in Sequence.* In §2 we considered a special case of variables x, y, z non-normally correlated in sequence, but possessing the property of linearity of regression, and found the coefficient of correlation between the measurement x of one set and the measurement z of another set, the two sets having the same measurement y, to be the product of the coefficients of correlation between x and y and between y and z. We now propose to show that linearity of regression is a sufficient condition that this be true in general. We take the three variables x, y, z correlated in sequence in accord with $F(x, y)$ and $G(y, z)$. We assume the functions to be continuous and the regression of y on x and of z on y to be

linear; that is,

$$\int yF(x, y)dy = (a_1x + b_1)f_1(x)$$

and

$$\int zG(y, z)dz = (a_2y + b_2)f_2(y),$$

where

$$f_1(x) = \int F(x, y)dy,$$

$$f_2(y) = \int F(x, y)dx = \int G(y, z)dz,$$

and a_1, a_2, b_1, b_2 are constants. Then

$$\psi(x, z) = \int \frac{1}{f_2(y)} F(x, y)G(y, z)dy.$$

We have

$$\frac{\int z\psi(x, z)dz}{\int \psi(x, z)dz} = \frac{\int\int z \frac{1}{f_2(y)} F(x, y)G(y, z)dzdy}{f_1(x)}$$

$$= a_1a_2x + a_2b_1 + b_2.$$

Thus the regression of z on x is linear and $\bar{r}_{zx}(\sigma_z/\sigma_x) = a_1a_2$. But $r_{xy}(\sigma_y/\sigma_x) = a_1$ and $r_{yz}(\sigma_z/\sigma_y) = a_2$. Therefore $\bar{r}_{zx} = r_{xy}r_{yz}$.

THE UNIVERSITY OF IOWA

ERROR EXPRESSIONS FOR CERTAIN CONTINUED FRACTIONS*

BY RAYMOND GARVER

Every discussion of the subject shows that the error committed in taking the nth convergent p_n/q_n as the true value of the simple continued fraction

$$(1)\quad a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\cdots+\cfrac{1}{a_n+\cdots}}}, \qquad (a_i \text{ positive integers}),$$

is less than $1/(q_nq_{n+1})$, and greater than $a_{n+2}/(q_nq_{n+2})$ in absolute value. Further, p_n/q_n is smaller than the true value if n is odd, and larger if n is even, provided the continued fraction does not terminate.

The purpose of the present paper is to supply alternate error limits for the important case when the continued fraction is the expansion of the square root of an integer N. These limits have the advantage that they do not require the computation of any convergent beyond the nth. Further, when applied as corrections to the nth convergent, they practically always lead to a much closer approximation than do the general error limits. Our results may be summarized as follows:

Given $N^{1/2}$ expressed in the form (1), with nth convergent p_n/q_n, and with k defined† as $|p_n^2-Nq_n^2|$. If n is even, the error committed in taking p_n/q_n as the value of $N^{1/2}$ is greater than $k/(2p_nq_n)$, and less than $k/(2q_n^2N)^{1/2}$, in absolute value. If n is odd, this statement is true with "greater than" and "less

* Presented to the Society, August 31, 1932.

† We might affix a subscript to k, but it does not seem necessary. If p_n and q_n are large the arithmetic required to compute k might be considerable if it were necessary to multiply out $p_n^2-Nq_n^2$. This is not, however, usually necessary. For, as is well known, the values that k can take on come in cycles, and the complete cycle can be formed easily from small values of n in many cases. It is also known that k is less than $2N^{1/2}$, which will render the complete squaring of p_n and q_n unnecessary even when the complete cycle of values has not been formed. Finally, if $n=mc$, where m is a positive integer and c is the number of partial quotients in the period of the continued fraction expansion of $N^{1/2}$, then k is 1. This case will receive special mention at the end of the present paper.

than" interchanged. These limits are ordinarily more satisfactory than the limits $1/(q_nq_{n+1})$ and $a_{n+2}/(q_nq_{n+2})$.

To obtain the desired error expressions we have merely to write

$$E \equiv \left| p_n/q_n - N^{1/2} \right| = k/(q_n(p_n + N^{1/2}q_n)).$$

Now if n is even, and we replace $N^{1/2}$ by p_n/q_n, which is larger than it, we have at once $E>k/(2p_nq_n)$. And if we, instead, replace p_n by $N^{1/2}q_n$, which is smaller than it, we find $E<k/(2q_n^2N^{1/2})$. To render this upper limit convenient for computation we may now replace $N^{1/2}$ by any approximation which is too small, say p_{n-1}/q_{n-1} or the better approximation Nq_n/p_n. If the $(n+1)$st convergent can be found easily it will be still better. The reader will see at once the modifications which must be made in these statements when n is odd.

It seems desirable to compare these error expressions with the general limits $1/(q_nq_{n+1})$ and $a_{n+2}/(q_nq_{n+2})$. First, the upper limit just determined is always less than $1/(q_nq_{n+1})$, when n is even. To see this note that the inequality

$$(2) \qquad \frac{k}{q_n(p_n + N^{1/2}q_n)} < \frac{1}{q_nq_{n+1}}$$

implies that

$$(3) \qquad kq_{n+1} < p_n + N^{1/2}q_n < 2p_n - \frac{a_{n+2}}{q_{n+2}}.$$

Since kq_{n+1} is an integer and $a_{n+2}<q_{n+2}$, we then have $kq_{n+1} \leqq 2p_n-1$. Now $k/(2q_n^2N^{1/2})$ is less than $1/(q_nq_{n+1})$ provided $kq_{n+1}<2N^{1/2}q_n$; and this inequality will be established as soon as we show that $2p_n-1<2N^{1/2}q_n$. This last is easily seen to be reducible to $E<1/(2q_n)$, which is satisfied since E is certainly less than $1/(q_nq_{n+1})$, and $q_{n+1}\geqq 2$ when n is even.

We have mentioned that $N^{1/2}$ may be replaced by suitable approximations to give a more convenient, though poorer, upper limit. Suppose we use p_{n-1}/q_{n-1}. A sufficient condition that the new upper limit still be less than $1/(q_nq_{n+1})$ is that $2p_n-1$ be less than $2q_np_{n-1}/q_{n-1}$. Using the relation $p_nq_{n-1}-q_np_{n-1} = (-1)^n$ this reduces to $2<q_{n-1}$. Hence the modified upper limit is less than $1/(q_nq_{n+1})$ unless q_{n-1} is 1 or 2; it may even be less in these trivial cases. The same statement holds, a fortiori, when

the modified upper limit is formed with a better approximation than p_{n-1}/q_{n-1}.

The lower limit $k/(2p_nq_n)$ for n even is greater than or equal to $a_{n+2}/(q_nq_{n+2})$ provided $a_{n+2} \leqq q_{n+1}$. This sufficient condition will certainly be satisfied unless n is small, for it is known that a_{n+2} is less than $2N^{1/2}$. To establish this relation between the lower limits note that the inequality

$$\frac{k}{q_n(p_n + N^{1/2}q_n)} > \frac{a_{n+2}}{q_nq_{n+2}} \tag{4}$$

implies that

$$kq_{n+2} > a_{n+2}(2p_n - Eq_n). \tag{5}$$

Now $k/(2p_nq_n) \geqq a_{n+2}/(q_nq_{n+2})$ provided $kq_{n+2} \geqq 2p_na_{n+2}$, and it follows from (5) that this condition will be satisfied whenever $a_{n+2}Eq_n \leqq 1$. But since $E < 1/(q_nq_{n+1})$, the sufficient condition as stated follows.

It is not easy to set up a simple example in which the new lower limit is poorer than the old. One can be found in connection with

$$31^{1/2} = 5 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{5 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{10 + \cdots}}}}}}}}. \tag{6}$$

If the second convergent, 6, be taken as the value of $31^{1/2}$, the error is greater than $5/12$ by the expression $k/(2p_nq_n)$, and greater than $3/7$ by the expression $a_{n+2}/(q_nq_{n+2})$. In this case, of course, we have $a_{n+2} > q_{n+1}$. If the fourth convergent be used instead of the second the two lower limits give the same value, but our new upper limit is better for the sixth convergent.

We turn to the consideration of the case where n is odd. An odd convergent is less in value than the continued fraction, so we have $p_n/q_n = N^{1/2} - E$, and in place of (3) we find

$$kq_{n+1} < 2p_n + Eq_n < 2p_n + \frac{1}{q_{n+1}}. \tag{7}$$

It then follows that $kq_{n+1} \leqq 2p_n$. And this is exactly the condition necessary to insure that the new upper limit $k/(2p_nq_n)$ is not greater than $1/(q_nq_{n+1})$.

When n is odd (5) must be written

$$kq_{n+2} > a_{n+2}(2p_n + Eq_n), \tag{8}$$

from which it follows that $kq_{n+2} \geqq 2p_n a_{n+2} + 1$. The lower limit for the case n odd, that is, $k/(2q_n^2 N^{1/2})$, will obviously be greater than $a_{n+2}/(q_n q_{n+2})$ provided $2p_n a_{n+2} + 1 > 2a_{n+2} q_n N^{1/2}$. And a moment's calculation, not greatly different from those already used, shows that this last inequality is certainly satisfied unless $2a_{n+2} > q_{n+1}$. This is again an exceptional condition, which can arise only for small values of n. Thus in $2^{1/2} = 1 + \frac{1}{2} + \frac{1}{2} + \cdots$, if we take the first convergent 1 as the value, we find the error greater than $1/2^{3/2}$ by the formula $k/(2q_n^2 N^{1/2})$, and greater than $2/5$ by the formula $a_{n+2}/(q_n q_{n+2})$. In this example the second lower limit is better than the first. But this is no longer true when n is 3 or larger.

It is again true that $N^{1/2}$ must be replaced by a rational approximation to get a lower limit suitable for computation; this time the approximation must be too large to insure that we still have a lower limit, but the values p_{n-1}/q_{n-1}, Nq_n/p_n, p_{n+1}/q_{n+1}, now have this property. And it is easy to show that the limit obtained with the aid of the first of these is larger than $a_{n+2}/(q_n q_{n+2})$ whenever $q_{n-1} > 2a_{n+2}$.

These general proofs that the new limits can only in exceptional cases be poorer than the old limits do not begin to show the advantage possessed by the new limits in most calculations. This can better be pointed out by taking the difference of the upper and lower limits in the two cases. If we replace $N^{1/2}$ in the one limit by the always available approximation Nq_n/p_n, we find that the difference of our two limits is $k^2/(2p_n q_n^3 N)$. This difference allows us to estimate at once how accurately we can approximate with the aid of the nth convergent. The corresponding difference between the usual limits $1/(q_n q_{n+1})$ and $a_{n+2}/(q_n q_{n+2})$ is $1/(q_{n+1} q_{n+2})$. The ratio of the first difference to the second is $R = k^2 q_{n+1} q_{n+2}/(2p_n q_n^3 N)$. Making obvious substitutions for q_{n+1} and q_{n+2} in terms of the earlier q's, and remembering that k, as well as each a_i, is less than $2N^{1/2}$, we find $R < 2(2N^{1/2} + 1)^2/(p_n q_n)$. The ratio R is then certainly less than 1 except for small values of p_n and q_n. In fact, R clearly approaches zero as n increases.

To have one numerical example, consider the approximation of $2^{1/2}$ by its sixth convergent, $99/70 = 1.414285714+$. The old limits simply show that this approximation is too large by an

amount larger than 0.0000700 and smaller than 0.0000846. However the new limits, with $N^{1/2}$ in the upper limit replaced by 140/99, tell us that the error is larger than 0.000072150 and smaller than 0.000072158. With this tremendous gain in accuracy we can at once say definitely that $2^{1/2}=1.41421356$, correct to 8 decimal places.

The limits of the present paper are applicable, under a certain restriction, to a well known method of approximating square roots by iteration. If a_1 is a first approximation to $N^{1/2}$, a second and closer approximation is obtained by taking $a_2=(a_1+N/a_1)/2$, a third approximation is formed similarly from the second, and so on. This method has been known since ancient times and recently has been dealt with in a number of papers.* Nevertheless it is interesting to point out that if c and m have the significance of the second footnote of this paper, and if a_1 is taken as p_{mc}/q_{mc}, that is, the mcth convergent to $N^{1/2}$, then a_2 is a later convergent, in fact, the $2mc$th.† Then a_3 is the $4mc$th convergent, and so on. There is, of course, ample justification for using the continued fraction expansion of $N^{1/2}$ to obtain a_1; we naturally want a close approximation without taking an unduly cumbersome fraction, and this is exactly what we obtain by using a convergent. And if we take the additional trouble to compute p_c/q_c, taking $m=1$ for convenience, we now know that the iterative method continues to give convergents of the form p_{mc}/q_{mc}. Our limits then apply, with $k=1$, and with n even, provided we form at least a_2. It is worthwhile mentioning that the lower limit $1/(2p_nq_n)$, if we write $a_i=p_n/q_n$, is equal to Bouton's lower limit $(a_{i-1}-a_i)^2/(2a_i)$, and is much easier to use. The proof of the equality is not difficult to make.

Similarly, the upper limit $1/(2q_n^2N^{1/2})$ is equal to Bouton's upper limit $a_i(a_{i-1}-a_i)^2/(2N)$ in case $N^{1/2}$ is replaced by Nq_n/p_n. Of course if a better approximation is available a better upper limit than Bouton's can be obtained.

THE UNIVERSITY OF CALIFORNIA AT LOS ANGELES

* Bouton, Annals of Mathematics, (2), vol. 10 (1908–9), pp. 167–72; James, American Mathematical Monthly, vol. 31 (1924), pp. 471–5.

† This follows from a theorem which may be found in Chrystal's *Algebra*, Part II, 1889, pp. 440–441. In a note appearing in the American Mathematical Monthly, vol. 39 (1932), pp. 533–535, I have gone into this point in more detail.

ON YOUNG'S DEFINITION OF AN ALGEBRA*

BY L. E. BUSH

1. *Introduction.* The usual definition of linear algebra† presupposes a number field over which the algebra is taken. Some of the properties of the algebra are dependent upon the properties of this associated field, for instance, the property that if e is an idempotent element, the subalgebra with basis e is a field and therefore permits unique division by every non-zero element.

Young‡ has given a definition of a general algebra and noted that a large part of the theory of linear algebras is valid for it. This definition does not introduce the associated field. Young concludes his paper with "Precisely to what extent, if at all, the present formulation is more general than the current one in terms of a field is a problem of no little interest." He notes the fact that every Dickson algebra is a Young algebra. That the converse of this is not true, however, is shown by the simple example of a modular ring. In fact, a modular ring possesses an idempotent element which generates by addition alone the entire algebra, yet division is not unique. Ingraham§ has noted this lack of restrictiveness of Young's definition, a lack of which, as I have been informed by Ingraham, Young had become cognizant.

The scalar multiplication in a Dickson algebra is not an operation within the algebra itself and it was with the view of eliminating this foreign operation that Young gave his definition. But, whether or not this operation is made a part of the definition of such an algebra, the possibility of defining an associated

---

* Presented to the Society, February 25, 1933.

† L. E. Dickson, *Algebren und ihre Zahlentheorie*, Zurich, 1927, pp. 23–24 and p. 32. A linear algebra as defined by Dickson will hereafter be referred to as a *Dickson algebra.*

‡ J. W. Young, Annals of Mathematics, (2), vol. 29 (1927–28), pp. 54–60. A general algebra as defined by Young will hereafter be referred to as a *Young algebra.*

§ M. H. Ingraham, this Bulletin, vol. 38 (1932), p. 100.

field and a scalar multiplication is certainly a property of every Dickson algebra, and the lack of this property in the Young algebra gives it its greater generality.

Assuming that it would be desirable to give a definition for a Dickson algebra which makes use only of operations within the algebra itself, it is the purpose of this paper so to extend Young's definition by the addition of postulates as to make it equivalent to Dickson's definition. It will be pointed out that these postulates, as applied to a Dickson algebra, are equivalent to certain of Dickson's postulates on scalar multiplication.

2. *Young's Definition.* In brief, Young's definition is as follows. Let $\mathfrak{S}$ be a system of two or more elements and two operations, addition and multiplication, and let the elements of $\mathfrak{S}$ form an abelian group, $\mathfrak{G}$, under addition. A class $[\mathfrak{H}]$ of subgroups of $\mathfrak{G}$ is called a *partition of* $\mathfrak{S}$ if every element a (except the identity) of $\mathfrak{G}$ is in one and only one subgroup $\mathfrak{H}_a$ of $[\mathfrak{H}]$. The partition $[\mathfrak{H}]$ is called an *L-partition* if the subgroup $\{\mathfrak{H}_1, \mathfrak{H}_2, \cdots, \mathfrak{H}_r\}$ of $\mathfrak{G}$ generated by any finite number r of subgroups $\mathfrak{H}$ is such that, if $\mathfrak{H}_k$ is any other $\mathfrak{H}$, then either $\{\mathfrak{H}_1, \mathfrak{H}_2, \cdots, \mathfrak{H}_r\}$ contains $\mathfrak{H}_k$ or it has no element in common with $\mathfrak{H}_k$ except the identity of $\mathfrak{G}$.

Let $[\mathfrak{H}]$ be an L-partition of $\mathfrak{S}$. Let (a) denote that $\mathfrak{H}$ which contains the element a and call it a *simple linear set of* $\mathfrak{S}$. By the *product* $(u)(v)$ of the simple linear sets (u) and (v) is understood the set of all elements of $\mathfrak{S}$ obtained by multiplying an element of (u) by an element of (v) in the order indicated. If for every choice of u and v, $(u)(v) = (uv)$, where (0) is the set consisting of the single element 0 (that is, the identity of $\mathfrak{G}$), $[\mathfrak{H}]$ is called an *invariant L-partition.*

Denote by $(a, b, \cdots, p)$ the subgroup of $\mathfrak{G}$ generated by $(a), (b), \cdots, (p)$ and call it a *linear set of* $\mathfrak{S}$. A finite set of elements $a_1, a_2, \cdots, a_r$ of $\mathfrak{S}$ is said to form *a basis of order r for* $\mathfrak{S}$ if there exists an invariant L-partition of $\mathfrak{S}$, such that (a_1), $(a_2), \cdots, (a_r)$ are all different and $(a_1, a_2, \cdots, a_r) = \mathfrak{G}$.

A Young algebra $\mathfrak{A}$ *of order r* (r finite) is a system consisting of a set $\mathfrak{S}$ of two or more elements, $a, b, \cdots$, and of two operations, addition $(+)$ and multiplication $(\cdot)$, and satisfying the following postulates. *If a, b, and c are any three elements of* $\mathfrak{S}$,

A1. $a+b$ *is a uniquely determined element of* $\mathfrak{S}$.

A2. $a+b=b+a$.

A3. $(a+b)+c=a+(b+c)$.

A4. *There exists in* $\mathfrak{S}$ *an element* 0 *such that* $a+0=a$.

A5. *There exists in* $\mathfrak{S}$ *an element* $-a$ *such that* $a+(-a)=0$.

From these postulates it follows that $\mathfrak{A}$ is an abelian group under addition and that the definitions in the first part of this section can therefore be applied to $\mathfrak{A}$.

M1. *ab is uniquely a determined element of* $\mathfrak{S}$.

M2. $a(b+c)=ab+ac$.

M3. $(b+c)a=ba+ca$.

F. $\mathfrak{A}$ *possesses a basis of order r* (*r finite*).

On the basis of these postulates Wedderburn's* calculus of complexes can be established and the first twenty-four theorems of his paper proved.

The definition of a Young algebra as given above is not exactly in the form in which Young stated it; however, only unessential changes have been made, particularly the explicit definition of the *order* of the algebra.

3. *The Associated System.* Let $\mathfrak{A}$ be a Young algebra. $\mathfrak{A}$ will be said to have *isomorphic linear sets with respect to its basis* if for every pair of elements a and b of $\mathfrak{A}$, neither of which is zero, there exists a simple isomorphism under addition between (a) and (b) such that a as an element of (a) corresponds to b as an element of (b).

Thus if $\mathfrak{A}$ has isomorphic linear sets there is an isomorphism between (a) and (b) in which any element a' of (a) corresponds to any element b' of (b). In particular, (a) is isomorphic with itself in such a manner that an arbitrary element a' of (a) corresponds to a. We shall introduce the abbreviation: a' *in* $(a)\sim b'$ *in* (b), which is to be read a' *as an element of* (a) *corresponds to* b' *as an element of* (b) *in that isomorphism between* (a) *and* (b) *in which* a *corresponds to* b.

$\mathfrak{A}$ will be said to have *equivalent linear sets with respect to its basis if*

(1) *it has isomorphic linear sets with respect to its basis;*

(2) a' *in* $(a)\sim b'$ *in* (b), c' *in* $(c)\sim d'$ *in* (d), $(ac\neq 0, bd\neq 0)$, *always implies* $a'c'$ *in* $(ac)\sim b'd'$ *in* (bd);

* J. H. M. Wedderburn, Proceedings of the London Mathematical Society, (2), vol. 6 (1908), pp. 79–81. Young, loc. cit., p. 60.

(3) a'' *in* $(a')\sim b'$ *in* (b), a' *in* $(a)\sim c'$ *in* (c), $(bc\neq 0)$, *always implies* a'' *in* $(a)\sim b'c'$ *in* (bc);
(4) b *in* $(a_1+a_2)\sim a_1'$ *in* $(a_1)\sim a_2'$ *in* (a_2) *always implies* $b=a_1'+a_2'$.

Let $\mathfrak{A}$ be a Young algebra with equivalent linear sets with respect to its basis and let $\mathfrak{A}^2\neq 0$. Let $\mathfrak{S}$ be a system consisting of a set of elements (which we shall denote by Greek letters) and of two operations, addition and multiplication. Let the elements of $\mathfrak{S}$ be in bi-unique correspondence with the elements of the simple linear set (a) of $\mathfrak{A}$, and let addition in $\mathfrak{S}$ be defined by postulating that $\mathfrak{S}$ be isomorphic with (a) under addition. Since $\mathfrak{A}$ has equivalent linear sets, $\mathfrak{S}$ is isomorphic under addition with every simple linear set of $\mathfrak{A}$. We define the product $\alpha\beta$ of any two elements α and β of $\mathfrak{S}$ as follows. Since $\mathfrak{A}^2\neq 0$, there exist in $\mathfrak{A}$ two elements a and b (which may be identical) such that $ab\neq 0$. Let a' in $(a)\sim\alpha$ in $\mathfrak{S}$ and b' in $(b)\sim\beta$ in $\mathfrak{S}$. The product $\alpha\beta$ is that element of $\mathfrak{S}$ which corresponds to the element $a'b'$ in (ab). That $\alpha\beta$ is thus uniquely defined is a consequence of the equivalence of the linear sets, for if c and d are two other elements of $\mathfrak{A}$ such that $cd\neq 0$ and these elements are used instead of a and b to define $\alpha\beta$, we have by (2) that $c'd'$ in $(cd)\sim a'b'$ in (ab) and are thus led to the same element $\alpha\beta$ of $\mathfrak{S}$. We shall call $\mathfrak{S}$ *the system associated with the basis of* $\mathfrak{A}$ *from which these linear sets arise.*

4. *Completion of the Definition.* In order to make Young's definition equivalent to that of Dickson we add to the postulates of §2 two others.

POSTULATE E. *Some basis of* $\mathfrak{A}$ *gives rise to equivalent linear sets.* Such a basis will be called an *E-basis of* $\mathfrak{A}$.

FIELD POSTULATE. *If* $\mathfrak{A}^2\neq 0$, *the system* $\mathfrak{S}$ *associated with some E-basis of* $\mathfrak{A}$ *is a field. If* $\mathfrak{A}^2=0$, *there exists a field simply isomorphic under addition with the linear sets of some E-basis of* $\mathfrak{A}$.

The two following theorems show the equivalence of the definitions. In the proof of the first theorem it becomes evident that, as applied to a Dickson algebra, the postulates here added are equivalent to Dickson's postulates on scalar multiplication.*

THEOREM 1. *Every Dickson algebra is a Young algebra which satisfies Postulate* E *and the Field Postulate.*

* Dickson, loc. cit., p. 24, Postulates II and III. The subdivisions of these postulates are referred to below as II (1), etc.

Let $\mathfrak{A}$ be a Dickson algebra of order n over a field $\mathfrak{F}$ with the basis $e_1, e_2, \cdots, e_n$. Then $\mathfrak{A}$ is a Young algebra and its basis as given above is a basis in the sense of Young. We have to show that Postulate E and the Field Postulate are satisfied.

If $x=\sum_{i=1}^{n} x_i e_i$ is any element of $\mathfrak{A}$, (x) consists of all elements αx as α varies over $\mathfrak{F}$. Let $y=\sum_{i=1}^{n} y_i e_i$ be another element of $\mathfrak{A}$ (y may be identical with x). Consider the correspondence between (x) and (y) in which $\alpha x \sim \alpha y$. In this correspondence $x \sim y$. The correspondence is evidently an isomorphism under addition and therefore satisfies (1). Since $(\alpha x)(\beta y)=(\alpha\beta)(xy)$,* (2) is satisfied. Since $\alpha(\beta x)=(\alpha\beta)x$,† (3) is satisfied. Since $\alpha(x+y)=\alpha x+\alpha y$,‡ (4) is satisfied. Thus Postulate E is satisfied by $\mathfrak{A}$ and since $\mathfrak{F}$ is a field the Field Postulate is also satisfied.

It may be noted that although Dickson's postulates II (1) and III (1) are not explicitly made use of in the above proof, they are implicitly assumed in the two added postulates, that is, in the possibility of relating a field to the algebra in the particular manner assumed.

LEMMA. *If $a_1, a_2, \cdots, a_n$ are linearly independent§ elements of a Young algebra which satisfies Postulate* E *and the Field Postulate and a_i' is in (a_i) $(i=1, 2, \cdots, n)$ and $\sum_{i=1}^{n} a_i'$ is in $(\sum_{i=1}^{n} a_i)$, then a_i' in $(a_i) \sim \sum_{i=1}^{n} a_i'$ in $(\sum_{i=1}^{n} a_i)$.*

For, let $c_k=\sum_{i=1}^{n} a_i - a_k$, so that $\sum_{i=1}^{n} a_i = a_k + c_k$, and let b_k' and c_k' be such that $\sum_{i=1}^{n} a_i'$ in $(a_k+c_k) \sim b_k'$ in $(a_k) \sim c_k'$ in (c_k). Hence, by (4), $\sum_{i=1}^{n} a_i' = b_k' + c_k'$. But c_k' is in (c_k), hence in the linear set $\mathfrak{C}=(a_1, a_2, \cdots, a_{k-1}, a_{k+1}, \cdots, a_n)$,‖ and $\sum_{i=1}^{n} a_i' - a_k'$ is for a similar reason in $\mathfrak{C}$. On the other hand b_k' and a_k' are in the linear set (a_k), and $(a_k) \wedge \mathfrak{C}=0$. Therefore $b_k'=a_k'$¶ and a_k' in $(a_k) \sim \sum_{i=1}^{n} a_i'$ in $(\sum_{i=1}^{n} a_i)$.

THEOREM 2. *Every Young algebra which satisfies Postulate* E *and the Field Postulate is equivalent to a Dickson algebra.***

---

* See Dickson, II, (3).

† See Dickson, II, (2).

‡ See Dickson, III, (2).

§ A set of elements is *linearly independent* in the sense of Young if no one of them belongs to the linear set generated by the others; Young, loc. cit., p. 57.

‖ Young, loc. cit., p. 57.

¶ Young, loc. cit., p. 56.

** Equivalent to a Dickson algebra with respect to the operations within

Let $\mathfrak{A}$ be a Young algebra of order n having the basis e_1, $e_2, \cdots, e_n$ satisfying Postulate E and the Field Postulate. If $\mathfrak{A}^2=0$, there is a field $\mathfrak{F}$ isomorphic under addition with the linear sets (e_1), (e_2), $\cdots$, (e_n). The zero Dickson algebra of order n over $\mathfrak{F}$ is evidently equivalent to $\mathfrak{A}$.

Suppose $\mathfrak{A}^2\neq 0$, and let $\mathfrak{S}$ be the system associated with the given basis. Then $\mathfrak{S}$ is a field. Since $\mathfrak{A}=(e_1, e_2, \cdots, e_n)$, $e_ie_j=\sum_{k=1}^{n}a_{ijk}$, where a_{ijk} is in (e_k). Let the elements of $\mathfrak{S}$ be denoted by Greek letters. Then a_{ijk} in $(e_k)\sim\alpha_{ijk}$ in $\mathfrak{S}$. Consider the Dickson algebra of order n over the field $\mathfrak{S}$ with the basis e_1', e_2', $\cdots$, e_n', where

$$e_i'e_j' = \sum_{k=1}^{n}\alpha_{ijk}e_k',$$

the α_{ijk} being defined as above. If x is any element of $\mathfrak{A}$, $x=\sum_{i=1}^{n}x_i$, where x_i is in (e_i). Let x_i in $(e_i)\sim\beta_i$ in $\mathfrak{S}$. Then to the element x of $\mathfrak{A}$ let correspond the element $x'=\sum_{i=1}^{n}\beta_ie_i'$ of the Dickson algebra $\mathfrak{A}'$.

Let x and y be any two elements of $\mathfrak{A}$ and x' and y' the corresponding elements of $\mathfrak{A}'$. Then

$$x = \sum_{i=1}^{n}x_i, \quad y = \sum_{i=1}^{n}y_i,$$

where x_i and y_i are in (e_i), and $x+y=\sum_{i=1}^{n}(x_i+y_i)$. Let x_i in $(e_i)\sim\beta_i$ in $\mathfrak{S}$, and y_i in $(e_i)\sim\gamma_i$ in $\mathfrak{S}$; then x_i+y_i in $(e_i)\sim\beta_i+\gamma_i$ in $\mathfrak{S}$. The element $x+y$ in $\mathfrak{A}\sim$ the element

$$\sum_{i=1}^{n}(\beta_i+\gamma_i)e_i' = \sum_{i=1}^{n}\beta_ie_i' + \sum_{i=1}^{n}\gamma_ie_i' = x'+y'$$

in $\mathfrak{A}'$. Therefore $\mathfrak{A}$ and $\mathfrak{A}'$ are isomorphic under addition.

Moreover,

$$xy = \sum_{i,j=1}^{n}x_iy_j, \quad x'y' = \sum_{i,j=1}^{n}\beta_i\gamma_je_i'e_j', \quad x_iy_j = \sum_{k=1}^{n}b_{ijk},$$

where b_{ijk} is in (e_k). Suppose $e_ie_j\neq 0$. Because of the invariance of the linear sets, x_iy_j is in (e_ie_j) and by the definition of multiplication in $\mathfrak{S}$, x_iy_j in $(e_ie_j)\sim\beta_i\gamma_j$ in $\mathfrak{S}$. Let a_{ijp_s}, $(s=1, 2, \cdots, r)$,

---

the algebra, that is, a Dickson algebra can be constructed which is isomorphic with the given algebra under addition and multiplication. Although scalar multiplication is not defined for the given algebra this theorem shows the possibility of defining such an operation for it.

be the set of all the a_{ijk}, $(k=1, 2, \cdots, n)$, which do not vanish. Then

$$e_ie_j = \sum_{k=1}^{n} a_{ijk} = \sum_{s=1}^{n} a_{ijp_s} \neq 0$$

and is contained in the linear set $\mathfrak{C}=(e_{p_1}, e_{p_2}, \cdots, e_{p_r})$. But $x_iy_j=\sum_{k=1}^{n} b_{ijk}$ is in (e_ie_j) and therefore also in $\mathfrak{C}$.* On the other hand, since b_{ijk} is in (e_k), $\sum_{s=1}^{r} b_{ijp_s}$ is in $\mathfrak{C}$ and it follows that

$$\sum_{k=1}^{n} b_{ijk} - \sum_{s=1}^{r} b_{ijp_s}$$

is in $\mathfrak{C}$. But the latter is evidently in a linear set supplementary to $\mathfrak{C}$ in $\mathfrak{A}$ and is therefore zero. All the elements appearing in this latter expression which do not vanish are, however, linearly independent; while a sum of linearly independent elements can not vanish.† Hence $b_{ijk}=0$, $k\neq p_s$, $(s=1, 2, \cdots, r)$. Similarly, since e_ie_j is in (x_iy_j), $b_{ijk}=0$ implies $a_{ijk}=0$. Such of the a_{ijk} as are different from zero are linearly independent and by the Lemma b_{ijk} in $(a_{ijk}) \sim x_iy_j$ in $(e_ie_j) \sim \beta_i\gamma_j$ in $\mathfrak{S}$. Because of the isomorphism of (e_k) and $\mathfrak{S}$ under addition, $\alpha_{ijk}=0$ if and only if $a_{ijk}=0$. By (3), b_{ijk} in $(e_k) \sim \beta_i\gamma_j\alpha_{ijk}$ in $\mathfrak{S}$, and b_{ijk} as an element of $\mathfrak{A}$ therefore corresponds to $\beta_i\gamma_j\alpha_{ijk}e_k'$ as an element of $\mathfrak{A}'$. If $e_ie_j=0$, $a_{ijk}=0$, $(k=1, 2, \cdots, n)$, and on account of the invariance of the linear sets $x_iy_j=0$, therefore $b_{ijk}=0$, $(k=1, 2, \cdots, n)$. But $a_{ijk}=0$ implies $\alpha_{ijk}=0$, and $b_{ijk}=0$ as an element of $\mathfrak{A}$ still corresponds to $\beta_i\gamma_j\alpha_{ijk}e_k'=0$ as an element of $\mathfrak{A}'$. Because of the isomorphism of $\mathfrak{A}$ and $\mathfrak{A}'$ under addition,

$$xy = \sum_{i,j=1}^{n} x_ix_j = \sum_{i,j,k=1}^{n} b_{ijk} \text{ in } A \sim \sum_{i,j,k=1}^{n} \beta_i\gamma_j\alpha_{ijk}e_k'$$

$$= \sum_{i,j=1}^{n} \beta_i\gamma_j e_i'e_j' = x'y' \text{ in } A'.$$

Thus $\mathfrak{A}$ and $\mathfrak{A}'$ are equivalent algebras.

THE OHIO STATE UNIVERSITY

---

* Young, loc. cit., p. 57.

† Young, loc. cit., p. 56.

SOME INVOLUTORIAL LINE TRANSFORMATIONS*

BY J. M. CLARKSON

1. *Introduction.* In a paper published in this Bulletin,† the author discussed three involutorial line transformations obtained by considering the transversals of one line each of four plane pencils; one line each of two plane pencils and two generators of a quadratic regulus; and two generators each of two quadratic reguli. In this paper we shall discuss those involutorial transformations between the two transversals of three generators of a cubic regulus and one line of a plane pencil; and the transversals of four generators of a quartic regulus. We shall also discuss an involutorial line transformation defined by two plane harmonic homologies. The methods used are algebraic and the results obtained are interpreted as point transformations on a certain quadratic hypersurface in S_5.

2. *Cubic Regulus and Plane Pencil.* Consider the ruled surface

$$F_3 \equiv z_1^2 z_3 - z_2^2 z_4 = 0,$$

and the pencil of lines in the plane

$$\alpha \equiv z_1 + z_2 - z_3 + z_4 = 0,$$

with vertex $A \equiv (1, 0, 1, 0)$. The Plücker coordinates of a generator of F_3 and those of a line of the pencil (A, α) are $(0, -k^2, 1, 0, k, k^3)$ and $(1, 0, -m, m, 1+m, 1)$, respectively, where the k and m are parameters. A general line (y) with coordinates y_i, $(i=1, \cdots, 6)$, meets three generators of F_3 and one line of (A, α). These four lines, in general mutually skew, have a second transversal (x) whose co-ordinates are found to be

$$x_i = \phi_i(y), \qquad (i = 1, \cdots, 6), \tag{1}$$

where the $\phi_i(y)$ are cubic functions in y_i. Thus the transformation is of third order.

The invariant locus of the transformation is a quadratic complex whose equation is found in the manner described in the pre-

* Presented to the Society, December 29, 1932.

† Vol. 38 (1932), pp. 533–540.

vious article by the author.* The singular elements of (1) are the two directrix lines of F_3, the bundle of lines (A), and the plane field of lines (α).

To the double directrix d_2 of F_3 correspond all of the lines of the special linear complex with axis a_{d_2}, the line of (A, α) met by d_2, and to the single directrix d_1 correspond the lines of the special linear complex whose axis is a_{d_1}. To each line of the bundle (A) corresponds the quadratic regulus determined by the three generators of F_3 met by the line of (A). Likewise, to each line of the plane field (α) corresponds the quadratic regulus determined by the three generators of F_3 met by the line of (α). The lines of (A) and of (α) belong also to the complex of invariant lines.

On the quadratic hypersurface V_2 in S_5 the point transformation (1) has for singular elements the points which represent the directrices d_1, d_2 of F_3 and the ρ-planes and ω-planes which represent the plane field (α) and the bundle (A), respectively.

To a general line on V_2 corresponds a cubic curve with the line as a bisecant. This can be seen from the fact that a line on V_2 represents a plane pencil in S_3 and a cubic curve on V_2 represents a cubic regulus in S_3. The two points of intersection represent the two invariant lines belonging to the pencil.

To a general plane of points on V_2 corresponds a cubic surface on V_2, for, in S_3, the image in (1) of a general bundle of lines is a congruence of order three, composed of the ∞^1 cubic reguli which are images of the ∞^1 pencils having their vertices at the vertex of the bundle; by duality, the image of a plane field of lines is a congruence of class three.

To the intersection of V_2 with a tangent hyperplane corresponds a three-dimensional locus of order three, for in S_3 the conjugate of a linear complex in (1) is a cubic complex, and the representation on V_2 of a special linear complex of S_3 is the intersection with V_2 of a hyperplane tangent to V_2 at the point which represents the axis of the complex.

3. *Quartic Regulus.* Let us now consider the involutorial line transformation determined by the transversals of four generators of a quartic regulus. Those ruled quartic surfaces having a straight line directrix are not considered, for this directrix is the

* Loc. cit.

second transversal of the four generators met by any arbitrary line of S_3. Hence only those quartics having a twisted cubic for a double curve and no simple directrix line are considered.

Let the double cubic be represented parametrically by

$$C_3 \equiv z_1 = t^3;\ z_2 = t^2;\ z_3 = t;\ z_4 = 1.$$

If a line join a point t of C_3 to a point s of C_3, its Plücker coordinates are

$$[1, -(t+s), ts, t^2s^2, ts(t-s), ts(t^2-s^2), (t^3-s^3)].$$

Now, a non-symmetric linear relation between t and s makes these coordinates of fourth degree in t, so that the ruled surface is a quartic. For the special case where $t=s$, the surface is the developable of C_3, and where $t+s=0$ we obtain the quadratic regulus common to three special linear complexes.

By choosing any non-symmetric relation between t and s, the transformation is found to be linear. It is the polarity as to the linear complex to which the given surface belongs. On V_2 in S_5, then, we have a projective transformation defined.

4. *The Harmonic Homologies.* Consider now in the plane $\alpha \equiv z_1 = 0$, the harmonic homology having its center at $A \equiv (0, 0, 0, 1)$ and as axis the line $a \equiv z_1 = 0,\ z_3 - z_4 = 0$. In the plane $\beta \equiv z_4 = 0$, consider the harmonic homology with center $B \equiv (1, 0, 0, 0)$ and axis $b \equiv z_4 = 0,\ z_2 - z_1 = 0$. A general line (y) in space meets α in a single point P and β in a point Q. In the homology $[a]$ in α, we have $P \sim P'$, and in the homology $[b]$ in β, we have $Q \sim Q'$. The line (y) then has for conjugate in the transformation the line $P'Q'$.

The equations of $[a]$ are

$$\rho z_1' = 0;\ \rho z_2' = z_2;\ \rho z_3' = z_3;\ \rho z_4' = 2z_3 - z_4,$$

while the equations of $[b]$ are

$$\rho z_1' = 2z_2 - z_1;\ \rho z_2' = z_2;\ \rho z_3' = z_3;\ \rho z_4' = 0.$$

Now an arbitrary line (y) with Plücker coordinates $(y_1, y_2, y_3, y_4, y_5, y_6)$ meets $z_1 = 0$ in the point $(0, y_4, y_5, y_6)$ and $z_4 = 0$ in $(y_6, -y_2, y_1, 0)$.

By $[a]$, the point $P \equiv (0, y_4, y_5, y_6) \sim P' \equiv (0, y_4, y_5, 2y_5 - y_6)$;

and, by $[b]$, the point $Q \equiv (y_6, -y_2, y_1, 0) \sim Q' \equiv (-y_6 - 2y_2, -y_2, y_1, 0)$. The Plücker coordinates of the line $P'Q'$ are

$$(2) \quad \begin{cases} x_1 = y_1(2y_5 - y_6), \ x_2 = y_2(2y_5 - y_6), \ x_3 = y_3y_6, \\ x_4 = y_4(-2y_2 - y_6), \ x_5 = y_5(-2y_2 - y_6), \\ x_6 = (2y_5 - y_6)(-2y_2 - y_6). \end{cases}$$

Then (2) are the equations of transformation.

The line AB is invariant. Its coordinates are (0, 0, 0, 0, 0, 1), which, when substituted in (2), make $x_6 \neq 0$ and $x_i = 0$, $(i = 1, \cdots, 5)$. Likewise, the pencils of lines having their vertices at A, B and lying in the planes A, b, and B, a, respectively, are such that each line is invariant, as can be determined by making the substitutions in (2). There is, in addition, a bilinear congruence of lines, namely, all lines meeting a, b, which are invariant. The equations of the congruence are $x_5 - x_6 = x_2 + x_6 = 0$.

Consider a line t_α in the plane α. It meets β in a point C of the line $c \equiv \alpha\beta$. In β there is a fixed point C_β, image by $[b]$ of C. In α, t_α has an image t_α' by $[a]$. Hence, since any point of t_α can be considered as P, the conjugate of t_α is the entire pencil (C_β, t_α'). Thus t_α is singular. As t_α revolves about C, t_α' describes a pencil whose vertex is C_α, the image by $[a]$ of C. Hence the conjugates of the lines of the pencil (C, α) belong to the bundle (C_β), all of whose pencils which are conjugates of t_α have their planes through $C_\alpha C_\beta$. As C describes the line c, C_β describes the line c_β in β. Thus the conjugate of the plane field of lines (α) is the special linear complex whose axis is c_β.

In like manner, the conjugate of the plane field of lines (β) is the special linear complex whose axis is c_α.

The intersection of the complexes $|c_\alpha|$, $|c_\beta|$ is a bilinear congruence which is the conjugate of the line c. Its equations are easily seen from (2) to be $2x_5 - x_6 = 2x_2 + x_6 = 0$.

Any line through a point C on c has for its conjugate the line $C_\alpha C_\beta$. Let the coordinates of C be $(0, 1, \lambda, 0)$ where λ is a parameter. The coordinates of C_α are $(0, 1, \lambda, 2\lambda)$ and of C_β $(-2, 1, \lambda, 0)$. Hence the coordinates of the line $C_\alpha C_\beta$ are quadratic in λ, so that to each line of a quadratic regulus $\{c\}_2$ corresponds by (2) a whole bundle of lines whose vertex lies on c. The regulus $\{c\}^2$ is singular.

If t_α revolve about a point $T \equiv (0, a, b, c)$, then t_α' describes

the pencil (T', α), where $T' \equiv (0, a, b, 2b-c)$, and C_β describes c_β. We have seen that t_α is singular and has as conjugate the pencil whose vertex is C_β and whose plane is (C_β, t_α'). Thus, the pencil (T, α) has as conjugate a congruence composed of the pencils whose vertices are on c_β and whose planes pass through T'. Let λ be a parameter. Then one representation of (T, α) will be the lines t_α joining $(0, a, b, c)$ to $(0, 0, 1, \lambda)$. Then t_α' will be the lines joining T' to $(0, 0, 1, 2-\lambda)$. The point C, where t_α meets c, is $(0, a\lambda, b\lambda-c, 0)$, and hence C_β is $(-2a\lambda, a\lambda, b\lambda-c, 0)$. The lines through an arbitrary point M and C_β have coordinates which are linear in λ. If any of these lines meet the corresponding t_α' they will belong to the congruence. Putting on the condition that two lines intersect, we obtain a quadratic equation in λ. Hence the order of the congruence is 2. An arbitrary plane meets c_β in one point C_β and meets the corresponding t_α' in one point. Hence the class of the congruence is 1.

On V_2 in S_5, a plane pencil of S_3 is represented by a line. Hence the conjugate by (2) of any line in the ρ-plane which is the representation of (α) is a ruled surface R on V_2 with a straight line directrix. The directrix is the representation of the pencil of S_3 whose vertex is T' and whose plane is (T', c_β). Every ω-plane of V_2 meets R in two points and every ρ-plane meets it in one. The S_3 determined by these two intersecting planes therefore meets R in three points; thus R is of order 3.

Consider the conjugate of the lines of an arbitrary plane pencil (T, τ). Each line of the pencil meets α and β in one point each, and hence has a definite conjugate. Since the conjugate of the pencil is the ∞^1 lines joining the corresponding points of two projective ranges on two skew lines, it is a quadratic regulus.

On V_2 in S_5, the conjugate of an arbitrary straight line (representation of a pencil in S_3) is a conic (representation of a quadratic regulus). This follows directly from (2) considered as a point transformation on V_2.

Consider now a plane field of lines in S_3. The plane μ of the field meets α and β in a line each, and the conjugate of (μ) is a (1, 1) congruence whose directrices are the images by $[a]$, $[b]$ of (μ, α), (μ, β), respectively. On V_2 in S_5 a general ρ-plane is transformed into a quadric surface.

The lines of S_3 belonging to a bundle (M), however, are trans-

formed by (2) into a congruence (3, 1). The fields (α), (β), projected from P upon, say, α are such that when P is joined to the self-corresponding points, then the line PA_i, $(i=1, 2, 3)$, passes through B_i, the point associated with A_i in the projectivity. Hence the congruence is of order three. A general plane π meets c_α, c_β in one point each, and the line joining π, c_α to π, c_β is the only line of the congruence lying in π. The class is therefore one.

On V_2 of S_5, a general ω-plane is transformed into a quartic surface.

The lines of S_3 belonging to a bilinear congruence with axes d_1, d_2 are transformed by (2) into a (4, 2) congruence. To see this, let us consider the quadratic regulus determined by d_1, d_2, c_β. These generators meet α in a conic α^2, whose image by $[a]$ is another conic α'^2, and meet β in a conic β^2, whose image by $[b]$ is a conic β'^2. The parameter λ of a point on α'^2 enters to the second degree; since the points of β'^2 are projective with those of α'^2, λ enters to degree two in defining the corresponding point on β'^2. The coordinates of the line joining an arbitrary point N to a point of α'^2 are of fourth degree in λ, when the line also passes through the corresponding point of β'^2. Thus there are four lines of the conjugate congruence which pass through N. An arbitrary plane ν meets α'^2 and β'^2 in two points each, and the lines joining the corresponding points are two in number. Hence the congruence is of class 2.

On V_2 of S_5, the conjugate of a quadric point cone (intersection with V_2 of the S_3 common to two tangent hyperplanes) is a surface of order six, for a general ρ-plane meets it in two points while a general ω-plane meets it in 4.

The conjugate of a special linear complex in S_3 is a quadratic complex, as is seen by the equations (2) which are of second degree. On V_2 of S_5, the intersection of a tangent hyperplane has for conjugate a three-dimensional locus of order two.

The author is now considering the case where there is in α, β, or both, some other Cremona involution.

CORNELL UNIVERSITY

ON A THEOREM OF HIGHER RECIPROCITY*

BY LEONARD CARLITZ

1. *Introduction.* Let $\mathfrak{D}$ denote the totality of polynomials in an indeterminate, x, with coefficients in a fixed Galois field of order p^n. Let P be a primary irreducible element of $\mathfrak{D}$; then, if A is any polynomial in $\mathfrak{D}$ not divisible by P,

$$A^{p^{n\nu}-1} \equiv 1 \pmod{P},$$

where ν is the degree of P. Evidently then

$$A^{(p^{n\nu}-1)/(p^n-1)}$$

is congruent (mod P) to a quantity in the $GF(p^n)$, that is, to a polynomial of degree zero. We define (A/P), the residue character of index p^n-1, as that element of $GF(p^n)$ for which

$$\left(\frac{A}{P}\right) \equiv A^{(p^{n\nu}-1)/(p^n-1)} \pmod{P}.$$

We have then the following theorem of reciprocity, proved in a recent paper.†

If P and Q are primary irreducible polynomials in $\mathfrak{D}$ of degree ν and ρ respectively, then

$$\left(\frac{P}{Q}\right) = (-1)^{\rho\nu}\left(\frac{Q}{P}\right). \tag{1}$$

The purpose of this note is to give a simple new proof of this theorem along the lines of Zeller's well known proof of the ordinary quadratic reciprocity theorem.‡

2. *Analog of Gauss' Lemma.* If A is in $\mathfrak{D}$, then sgn A denotes the coefficient of the highest power of x which occurs in A; if sgn $A=1$, A is *primary*. Let $\mathfrak{R}(A/B)$ denote the remainder in the division of A by B. Then the analog in question is furnished by the following theorem.

---

* Presented to the Society, December 27, 1932.

† American Journal of Mathematics, vol. 54 (1932), pp. 39–50.

‡ Monatsbericht der Berliner Akademie, December, 1872.

LEMMA. *Let A and P be in $\mathfrak{D}$, P primary irreducible, and not a divisor of A; then*

$$\left(\frac{A}{P}\right) = \prod_H \operatorname{sgn} \mathfrak{R}\left(\frac{HA}{P}\right), \tag{2}$$

the product extending over all primary H of degree less than the degree of P.

A detailed proof of this lemma is scarcely necessary, but we remark for a later purpose that the proof depends on the fact that the set of polynomials

$$\left\{ \mathfrak{R}\left(\frac{HA}{P}\right) \Big/ \operatorname{sgn} \mathfrak{R}\left(\frac{HA}{P}\right) \right\}$$

is identical (except for order) with the set $\{H\}$, where H runs through the primary polynomials of degree less than the degree of P.

3. *Proof of the Theorem.* Let $\{bM\}$ denote the set of those polynomials in the set

$$\left\{ \mathfrak{R}\left(\frac{HQ}{P}\right) \right\}, \qquad (\deg H < \nu, \operatorname{sgn} H = 1),$$

with signum equal to b, a fixed quantity in $GF(p^n)$. We write $S_b = \{M\}$; evidently the polynomials M are primary. Similarly we put $S_b' = \{N\}$, where $\{bN\}$ denotes the set of those polynomials in the set

$$\left\{ \mathfrak{R}\left(\frac{KP}{Q}\right) \right\}, \qquad (\deg K < \rho, \operatorname{sgn} K = 1),$$

with signum equal to b.

We assume, as we may without any loss in generality, that $\rho \geqq \nu$. We put

$$S_b' = U_b + V_b, \tag{3}$$

where

$$U_b = \{M \text{ in } S_b'; \deg M < \nu\},$$
$$V_b = \{M \text{ in } S_b'; \deg M \geqq \nu\}.$$

Then we begin by proving

$$U_b = S_{-b}. \tag{4}$$

Indeed, let M be any polynomial in S_b, that is, let

$$HQ \equiv bM \pmod{P},$$

where deg $H<\nu$, sgn $H=1$. Evidently there exists a primary K of degree $<\rho$ such that

$$HQ = bM + KP.$$

But this equality may be written in the form

$$KP \equiv -bM \pmod{Q},$$

and since deg $M<\nu$, it follows at once that M is in U_{-b}.

Conversely assume an M in U_b. Then there exists a primary K of degree $<\rho$, such that

$$KP \equiv bM \pmod{Q};$$

since deg $M<\nu$, we infer the existence of a primary (in particular, non-zero) H of degree $<\nu$, such that

$$KP = bM + HQ.$$

Then it follows as above that M is in S_{-b}. We have therefore set up a (1, 1) correspondence between the elements of U_b and S_{-b}, thus proving equation (4).

Let us write $\mu(W)$ for the number of elements in a (finite) set W. Then, by Gauss' Lemma and equation (3),

$$\left(\frac{P}{Q}\right) = \prod_{b\neq 0} b^{\mu(U_b)+\mu(V_b)}, \tag{5}$$

the product in the right member extending over all b in $GF(p^n)$ different from zero. By equation (4), the right side of (5) may be written in the form

$$\prod_b b^{\mu(S_{-b})+\mu(V_b)}.$$

Now

$$\prod_b b^{\mu(S_{-b})} = \prod_b (-b)^{\mu(S_b)} = \prod_b (-1)^{\mu(S_b)} \cdot \prod_b b^{\mu(S_b)}; \tag{6}$$

but (by the remark in §2)

$$\sum_b \mu(S_b) = \frac{p^{n\nu}-1}{p^n-1} \equiv \nu \pmod 2, \tag{7}$$

and by Gauss' Lemma

$$\prod_b b^{\mu(S_b)} = \left(\frac{Q}{P}\right); \tag{8}$$

we have therefore, by equations (5), $\cdots$, (8),

$$\left(\frac{P}{Q}\right) = (-1)^\nu \left(\frac{Q}{P}\right) \prod_b b^{\mu(V_b)}. \tag{9}$$

It remains to calculate $\mu(V_b)$; evidently we may ignore the case $b=1$.

Let M be in V_b, so that for some primary K of degree $<\rho$,

$$KP \equiv bM \pmod Q. \tag{10}$$

Since the degree of M is not less than the degree of P, we may put

$$M = AP + cB, \quad \deg B < \nu,$$

where A and B are primary, and c is in $GF(p^n)$. Then (10) becomes

$$(K - bA)P \equiv cB \pmod Q. \tag{11}$$

But

$$\deg A = \deg M - \deg P < \rho - \nu,$$

and since we are assuming $b \neq 1$, we have necessarily

$$\deg K \geqq \rho - \nu.$$

Therefore $K - bA$ is primary and

$$\rho - \nu \leqq \deg (K - bA) < \rho,$$

and finally B is in U_c.

Conversely, let us begin with a B in U_c:

$$KP \equiv cB \pmod Q, \quad \deg B < \nu. \tag{12}$$

Then, if m is an integer such that

$$\nu \leqq m < \rho, \tag{13}$$

and A is a primary polynomial of degree $m-\nu$, we have

$$(K + bA)P \equiv bAP + cB \pmod{Q};$$

and if we put

$$bAP + cB = bM,$$

it is evident that M is in V_b. Indeed

$$\begin{aligned} \deg (K + bA) &= \deg K, \\ \operatorname{sgn} (K + bA) &= \operatorname{sgn} K = 1; \\ \deg M &= \deg AP = m, \\ \operatorname{sgn} M &= b^{-1} \operatorname{sgn} (bAP + cB) = 1. \end{aligned}$$

To sum up, we have proved that, for fixed $b \neq 1$,

(i) to each element of V_b corresponds a single element of some U_c;

(ii) to each element of U_c, c fixed, corresponds $p^{n(m-\nu)}$ elements of V_b of degree m, where m is a fixed integer satisfying the inequalities (13).

Evidently (ii) implies that the total number of elements of V_b corresponding to a fixed element of U_c is precisely

$$\sum_{m=\nu}^{\rho-1} p^{n(m-\nu)} = \frac{p^{n(\rho-\nu)} - 1}{p^n - 1}.$$

We have therefore that the number of elements in V_b is

$$\mu(V_b) = \frac{p^{n(\rho-\nu)} - 1}{p^n - 1} \frac{p^{n\nu} - 1}{p^n - 1}, \tag{14}$$

as follows at once from equations (4) and (7).

Returning to equation (9), we have, since the right member of (14) is independent of b,

$$\prod_b b^{\mu(V_b)} = \left(\prod_b b \right)^{\epsilon}, \tag{15}$$

ϵ denoting the right side of (14).

Now, by the generalization of Wilson's Theorem for a Galois field,

$$\prod_b b = -1;$$

on the other hand

$$\epsilon \equiv (\rho - \nu)\nu \equiv \rho\nu + \nu \pmod 2;$$

therefore, by (9) and (15),

$$\left(\frac{P}{Q}\right) = (-1)^{\rho\nu}\left(\frac{Q}{P}\right).$$

This completes the proof of our theorem of higher reciprocity.

4. *Remarks.* It should be clear that the case $p=2$ is by no means ruled out in the proof just given. Since in the $GF(2^n)$, $+1$ and -1 are the same, the theorem in this case assumes the simpler form

$$\left(\frac{P}{Q}\right) = \left(\frac{Q}{P}\right), \qquad (p = 2),$$

(P/Q) being the residue character of index 2^n-1.

Secondly, if in the notation of §3, we put

$$W_b = \{M \text{ in } S_b'\,;\ \deg M \leqq \nu\}$$

and

$$X_b = \{M \text{ in } S_b'\,;\ \deg M > \nu\},$$

then it is easy to show that

$$\prod_b b^{\mu(X_b)} = (-1)^{\rho\nu},$$

or, what amounts to the same thing,

$$\prod_b b^{\mu(W_b)} = \left(\frac{Q}{P}\right).$$

DUKE UNIVERSITY

OFFICIAL COMMUNICATIONS

Meetings of the Society have been fixed at the following times and places:

STANFORD UNIVERSITY, March 18, 1933.

Abstracts must be in the hands of Associate Secretary T. M. Putnam, University of California, Berkeley, California, not later than February 25. Abstracts received by that date will appear in the March issue of the BULLETIN.

NEW YORK CITY, April 14-15, 1933.

Abstracts must be in the Office of the Society, 501 West 116th St., New York City, not later than March 24. Abstracts received by March 6 will appear in the March issue of the BULLETIN. A symposium on algebra is being arranged. The leading speakers will be Professor Oystein Ore, on *Abstract ideal theory and applications*, and Professor J. V. Uspensky, *On the theory of numbers and integral calculus*.

CHICAGO, April 14-15, 1933.

Abstracts must be in the hands of the Associate Secretary, Mark H. Ingraham, University of Wisconsin, Madison, Wisconsin, not later than March 25. Abstracts received by March 6 will appear in the March issue of the BULLETIN. Professor C. C. MacDuffee will deliver a symposium lecture on *Matrices with elements in a principal ideal ring*. There will be a special memorial meeting in honor of Professor E. H. Moore, at which several of his students will speak.

THIRTY-NINTH SUMMER MEETING, CHICAGO, June 19-24, 1933.

This meeting will be in conjunction with the meeting of the American Association for the Advancement of Science at the Century of Progress Exposition, but most of the actual sessions will be at the University of Chicago. Addresses will be delivered by Professors Lipót Fejér of the University of Budapest, Tullio Levi-Civita of the University of Rome, George D. Birkhoff of Harvard University, and Leonard E. Dickson of the University of Chicago. Other addresses and symposia in which American and foreign speakers will participate are being planned. The Mathematical Association of America will meet during the same week. Titles and abstracts of papers to be presented at this meeting should be in the hands of Associate Secretary Mark H. Ingraham, University of Wisconsin, Madison, Wisconsin, not later than May 20. Abstracts received before May 6 will appear in the May issue of this BULLETIN.

R. G. D. RICHARDSON, *Secretary of the Society.*

Articles for insertion in the BULLETIN should be addressed to E. R. HEDRICK, Editor of the BULLETIN, University of California at Los Angeles. Reviews should be sent to W. R. LONGLEY, Yale University, New Haven, Conn. Notes should be sent to H. W. KUHN, Ohio State University, Columbus, Ohio.

Subscriptions to the BULLETIN, orders for back numbers, and inquiries in regard to non-delivery of current numbers should be addressed to the American Mathematical Society, 450-459 Ahnaip St., Menasha, Wis., or 501 West 116th St., New York.

The initiation fees and the annual dues of members of the Society (see this BULLETIN, p. 322, May, 1930; and the List of Officers and Members, September, 1932, p. 66), are payable to the Treasurer of the Society, Professor G. W. Mullins, 501 West 116th St., New York City.

WHOLE NUMBER 401

CONTENTS

PAGE

Thermodynamics and Relativity. By R. C. TOLMAN.... 49

Non-Analytic Functions of a Complex Variable. By E. R. HEDRICK..... 75

Sets of Local Separating Points of a Continuum. By G. T. WHYBURN..... 97

On the Existence of the Absolute Minimum in Problems of Lagrange. By L. M. GRAVES..... 101

Some New Factorizations of $2^n \pm 1$. By D. H. LEHMER... 105

Matrices whose sth Compounds are Equal. By JOHN WILLIAMSON..... 108

Remarks on Propositions $*1{\cdot}1$ and $*3{\cdot}35$ of Principia Mathematica. By B. A. BERNSTEIN..... 111

On the Class Numbers of a Cyclic Field and a Sub-Field. By C. G. LATIMER..... 115

Some Applications of Murphy's Theorem. By H. BATEMAN..... 118

On singular Chains and Cycles. By S. LEFSCHETZ..... 124

Variables Correlated in Sequence. By A. T. CRAIG..... 129

Error Expressions for Certain Continued Fractions. By RAYMOND GARVER..... 137

On Young's Definition of an Algebra. By L. E. BUSH.... 142

Some Involutorial Line Transformations. By J. M. CLARKSON..... 149

On a Theorem of Higher Reciprocity. By LEONARD CARLITZ..... 155

For official Communications and Notices, see the inside of the back cover.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN